System Simulation Ltd

The Significant Properties of Moving Images

JISC

Project	JISC Digital Preservation Programme : Study on the Significant Properties of Moving Images
Project Acronym	SP-MI
Authors	Mike Coyne Mike Stapleton
Contact name	Mike Stapleton
Document name	The Significant Properties of Moving Images
Version:	6.4
Date:	26 March 2008

Acknowledgements

Many thanks to those who contributed to this Study by participating in our workshops and online discussions and commenting on our various conclusions: Kevin Ashley, David Cruickshank, David Duce, Bob Hopgood, James Insell, Andrea Kallas, Gareth Knight, Pip Laurenson, George Mallen, Brian Matthews, C Nielson and Nick Townend.

We should especially like to thank Richard Wright for his particular contribution to the project.

The project was funded under the JISC Digital Preservation Programme and the copyright of this document is the property of JISC. The moral rights of the study remain with the authors.

Table of Contents

Executive Summary

This chapter needs revision to reflect the current version of the text

This study is funded by the Joint Information Systems Committee of the UK Higher Education funding bodies (JISC), under the Digital Preservation and Records Management Programme, and is part of a series of parallel studies that are aiming to contribute to a broader strategy for digital preservation by characterising the significant properties of a series of digital objects.

The aim of this particular study is to contribute to the development of a consistent methodology, especially for the preservation of digital moving images stored by archives in the UK Higher Education sector.

The study is relatively modest in its ambitions. It builds on a substantial body of earlier work (notably that brought together in the 2006 Digital Moving Images and Sound Archiving Study, undertaken for JISC by the Arts & Humanities Data Service (AHDS)) and aims to contribute further to the development of practical procedures for professionals engaged in digital preservation. In particular, it proposes a schema, based on the model developed by the InSPECT project, for characterising the Significant Properties of digital moving images.

The substance of the study is as follows:

The Dilemma of Digital Preservation

There is now a considerable amount of moving image material in digital form – both digitised film and video and born-digital film and video. But, equipment and computer programs change over time and the ability to read older files and display the information may be lost. Obsolescence is a major problem and digitisation as such does not solve the problems of preservation.

The general response to these issues is to plan for the longer term, to decide on preservation formats that are likely to be relatively long lived, but also to plan for the inevitable changes over time. This can be done either by *migrating* data successively on to new formats or by developing software and equipment that enable older programs to be *emulated* in future environments.

This is where the core issue of this study arises. If the original object necessarily has to be changed in the course of processes designed to ensure its preservation over the longer term, which are the characteristics and features of the original that have to be preserved through these changes? That is, what are the *Significant Properties* of the object?

The next question is how are these Significant Properties to be preserved? If migration is adopted as a strategy, there is a need to check how far a preservation format captures the Significant Properties of the original. If an emulation strategy is chosen, the emulation software has to be checked to see if it is successfully reproducing the Significant Properties of the original.

If the required properties are not all captured in the preservation format they must be recorded as supplementary metadata in an appropriate schema. The metadata schema should also accommodate wider factors bearing on the authentic reproduction of digital objects and notably the 'intent' of the original creator.

A Framework for Digital Preservation

The Open Archival Information System (OAIS) Reference Model provides a general framework for digital preservation. Archives conforming to the OAIS model accept the responsibility to preserve information over the *long term* and to make it available for a *Designated Community*.

One of the objectives in developing the Reference Model was to facilitate the effective participation of non-archival organisations in the preservation process, particularly by developing a vocabulary that can be shared by the range of stakeholders. Further clarification of terminology and usage in the preservation area is provided by the PREMIS (PREservation Metadata: Implementation Strategies) Data Dictionary.

The concept of the Archival Information Package (AIP) is of particular interest. Its place within the Reference Model is explained in more detail in the body of the Study, but in an AIP we would expect to find :

- the data object encoding the content (a set of bit sequences);
- information about how to make these data meaningful to users;
- a further set of information that is necessary for adequate preservation of the target information, primarily relating to Provenance, Reference, Fixity, and Context information

Our main task in this study is to determine significant parts of the information set that is necessary in order to meet the second requirement of the AIP and to continue to meet this over time.

Meeting the third requirement is usually seen as being a matter of providing an appropriate metadata *wrapper*. Metadata are information about data. The respective merits of MPEG-21 and METS are briefly considered

It is noted that the EU research project CASPAR is currently developing both conceptual aspects of the OAIS reference model and practical tools to help its implementation particularly in relation to cultural and scientific heritage.

A Framework for Determining Significant Properties

The immediate framework within which we are setting out the significant properties of digital moving images is the JISC-funded project : Investigating the Significant Properties of Electronic Content Over Time (InSPECT)

InSPECT adopts the distinctive *Performance Model* initially developed by the National Archives of Australia. This sees each interaction between a data source and the technology that presents it as a *performance*, by analogy with the way that a play or a piece of music is performed. In the Performance Model, the role of Significant Property analysis is to define those properties that it is necessary to reproduce in order to provide an 'authentic' performance.

Within this framework InSPECT to propose a taxonomy of properties expressed in terms of *Content, Context, Rendering, Structure* and *Behaviour.*

We can see here a further development of the OAIS framework. Before beginning to apply this framework to moving images, it is noted that the European PLANETS project, (with a significant input from The National Archives) is planning to develop methodologies and tools for evaluating preservation actions and their ability to recognize and reproduce Significant Properties and to develop further the tools for automatically recognizing and processing specific digital objects, made available by the PRONOM technical registry.

Types of Moving Images & their Digitisation

Moving images consist broadly of a sequence of still images projected at a rate that the eye and brain interprets as continuous motion. In the case of video the story is slightly more complex. The study provides an overview of the main types of moving images in analogue and digital form and this informs the identification of their Significant Properties.

The results of digitisation are an important consideration. Indeed, a major conclusion of the study is that a large proportion of the core, defining, technical characteristics of moving image sequences - their Significant Properties - are actually captured in the digitisation process or by the formats used for born-digital moving images. This implies that ensuring that the Significant Properties of moving images are reproduced is, to an important extent, a matter of using the right format.

Bringing together elements from the Digital Moving Images and Sound Archiving Study, the InSPECT methodological approach and the results of our own discussions and analysis, the study proposes that the core technical Significant Properties are:

Content	Number of image and audio streams; length;
Context	Title; creator; date of creation; provenance;
Rendering	Frame width and height; bit-depth; gamut; pixel aspect ratio; frame rate; compression ratio; codec
Structure	Relationship between audio and image streams, relationship between metadata and bit-streams; interlace;

The choice of the preservation format is critical in capturing and preserving these Significant Properties. There are a number of criteria for making this choice and those of the US Library of Congress are widely used as a reference point.

The Library of Congress also proposes a particular set of moving image file formats on the basis that they are not copy-protected and that they allow relatively complete descriptive and technical metadata. It does, however, comment: ‘The formats in this category are emerging, and this list of acceptable formats for *uncompressed or lossless compressed* data should be treated as provisional’.

The conclusion at this stage is that capturing most of the Significant Properties of Moving Images can be achieved by using a relatively small range of formats for preservation purposes and the advantage of the approach at this level is that most of the core Significant Properties are captured automatically or at least routinely. There is, however, another level at which information about digital objects can be recorded, and although this usually requires a direct (manual) input of data at some stage and therefore generates corresponding costs, it does allow a much greater range of information to be recorded.

Content & Intent in Moving Images

Although the focus so far has been elsewhere, the circumstances of the creation of moving images and the intentions of the creator are of concern in the determination of Significant Properties. Our discussions with film archivists suggested that this can be an issue of considerable importance in the preservation of moving images.

The InSPECT concept of *performance* suggests the associated notion of *interpretation* and the consideration that different art forms expect and accommodate differing degrees of interpretation.

We take it as read that a significant proportion of moving image archivists and of the user community regards the original intentions of the moving image creators as important and, while many features of are reducible to different technical characteristics, there are some contextual considerations that cannot captured automatically and these have to be expressed through appropriate metadata.

Conclusions & Recommendations

The main technical conclusions of the Study have already been set out in this summary. However, a series of more general conclusions have also been drawn.

Our focus has been on the main types of digital moving image (held in file format) that of interest to digital archives. This includes digitised film, digital cinema and broadcast and other output held in digital video. Unfortunately, we were not able to give much consideration to medical and scientific imaging and our consideration of digital video has been restricted to device-independent video. These exclusions indicate areas for further investigation.

We believe that our Study has built successfully on the OAIS Reference Model and in particular elaborated the concept of an Archival Information Package (AIP) for moving images. We point out that it may also inform the choice of the Dissemination Information Package (DIP) – the package that allows users access to the digital moving images held by the archive. However complementary work needs to be undertaken in relation to the Submission Information Package (SIP) – how objects for preservation are delivered to the archive by the producer. The development, in particular, of a comprehensive metadata schema for the archiving of digital moving images requires co-ordination with work undertaken in these complementary areas.

In developing an appropriate metadata schema, for the explicit recording of Significant Properties, we have supported the developing use of METS and suggested how METS might be further developed to accommodate the requirements of digital archiving, as we have outlined them. We believe that a flexible regime, such as that provided by METS, is necessary as a way of taking account of the broader type of Significant Property relating to stylistic and aesthetic features and especially those that arise from the intent of the creator of the moving image sequence.

Recommendations

There are a series of other recommendations made on the definition of Significant Properties in related areas:

- Digital video not stored in files;
- Interactives;
- Medical and scientific imaging.

The full development of the metadata schema for the archiving of digital moving images will require resolution of issues that are beyond the scope of this study. A recommendation is made that a general solution be found.

Finally we note that it is important that developments in digital preservation in the UK relate effectively to those taking place elsewhere and not only in the English-speaking world. It will be particularly important to take up the results of the CAPAR and PLANETS projects, not least through the agency of the UK institutions that are active in them and we recommend that a watching brief be kept on them.

1 Introduction

1.1 The Aims of the Study

This study is funded by the Joint Information Systems Committee of the UK Higher Education funding bodies (JISC)[1], under the Digital Preservation and Records Management Programme[2], and is part of a series of parallel studies that are aiming to contribute to a broader strategy for digital preservation by characterising the significant properties of a series of digital objects

The aim of this particular study is to contribute to the development of a consistent methodology for the preservation of moving images in digital form, especially digital moving images stored by archives in the UK Higher Education sector, but its remit requires that a wider context be taken into account, extending both to work on other digital objects and to other significant developments in preservation strategy in the UK, across Europe and beyond.

It will be seen that the study is able to build on a substantial body of earlier work, covering a wide range of issues. However, the challenges are equally wide. There is a huge amount of moving image material, mostly still in analogue form, that ought to be preserved. A series of digitisation programmes are under way, but often they are struggling to keep up with the problem and on occasions they are inevitably cutting corners. The technology continues to evolve rapidly and, given that preservation requires taking a view on likely developments over a long time scale, judgements about the appropriate forms are intrinsically difficult to make. Even the material that has been digitised already or is born digital has a variety of forms and in most cases the digitisation has been undertaken with current use and access in mind rather than with a view to longer-term preservation. Practical experience in the digital preservation of moving images is therefore limited and whereas archives are beginning to develop professional practice and norms in the area, most see themselves as being at the beginning of this process.

Furthermore, there continue to be clear differences of view about the nature and function of preservation itself. There are a number of different perspectives on the requirements of future users of preserved digital material and differences of view are evident concerning the significance of the intentions of the original creators.

In these circumstances, the current study has to be relatively modest in its ambitions. The aim is to contribute further to the development of practical procedures for professionals engaged in digital preservation and specifically to propose a schema for the characterisation of the significant properties - for preservation purposes - of digital moving images, but in so doing, it will have to base itself in those areas, where a consensus is established, or at least clearly emerging, and to restrict itself to suggesting tentative extensions of current thinking in other areas where the way forward is by no means clear. It is hoped, however, that the study will serve a practical purpose in providing a clear

1 http://www.jisc.ac.uk/

2 http://www.jisc.ac.uk/whatwedo/programmes/programme_preservation

statement on the areas of current consensus and in directing attention to the practical steps to be taken by archivists in their important task of preserving artefacts that are a central cultural feature of modern societies.

To that end the study will set out rather schematically the issues to be addressed in the long term preservation of digital moving images, hopefully making the practical implications for archivists abundantly clear. We also aim to make the document relatively accessible to people who confront preservation issues, but who do not have the benefit of extensive technical knowledge in the area. Frequent reference will be made to earlier and parallel work, particularly where this can provide further explanation of the position taken. Notable among these sources is the Digital Moving Images and Sound Archiving Study, undertaken for JISC by the Arts & Humanities Data Service (AHDS) and completed in 2006[3]. This provides an immediate reference point for our own work. Equally, we hope we have been able to add fresh insights, not least those derived from the very helpful inputs to a series of workshops conducted over the summer and autumn of 2007.

1.2 The Structure of the Study

After an initial description of the issues that the Study is intended to address, the discussion first seeks to establish the framework within which the definition of the Significant Properties of moving images is to take place. This requires briefly considering aspects of the OAIS Reference Model that is widely regarded as providing a general framework for digital archives and, in particular, an outline of the concept of an Archival Information Package, which is the primary point of focus for our later discussion. Similarly, there is a brief reference to the metadata schemata that are emerging as the main contenders in digital archives.

The Study then goes on to look specifically at the framework for defining Significant Properties provided by the most recent work under the InSPECT project. InSPECT is developing a general approach to the characterisation of Significant Properties and thus provides a basic orientation for our own approach.

Reference is also made to work being undertaken by CASPAR and PLANETS, projects funded by the European Union that are both developing existing conceptual frameworks and making available certain tools that will be of practical use to digital archives. Co-ordination with European and global frameworks is of growing importance for these initiatives.

This explanation of the framework for defining Significant Properties is then followed by an analysis of the properties of moving images. The main types of moving image are established, and this is followed by some observations of the process of digitisation and the consequences for the study of the digital forms that moving images can assume. This then allows the study to proceed to a

---

[3] 'Digital Moving Images and Sound Archiving Study' AHDS 2006

http://www.jisc.ac.uk/media/documents/programmes/preservation/moving_images_and_sound_archiving_study1.pdf characterisation of the core Significant Properties of moving images and a determination of the formats that can currently be adopted for preservation purposes.

After a determination of these central issues, there is consideration given to a broader range of Significant Properties and especially to those that arise from the original intent of the creators of this kind of digital object. A metadata framework is suggested that can accommodate this type of Significant Property.

Finally, the main conclusions are drawn together and recommendations are made for future developments.

1.3 The Dilemma of Digital Preservation

There is a vast amount of film and video that is worth preserving but exists only in physical formats such analogue film and magnetic tape. Given an appropriate environment, film can be preserved in the long term, however, in spite of such care, the magnetic tapes are deteriorating, increasingly to the extent that either they can no longer be used or it cannot be used without extraordinary care and effort. Digitisation is central to preservation strategies as well as to plans to increase access. It has become the main focus of public policy in the area and a vehicle for achieving greater profile in the wider policy arena. The European Union's Digital Libraries initiative provides an illustration of this developing policy agenda. Increasingly new content is created in a digital medium ('born digital'), adding to the demands for digital preservation.

Preserving material in digital form poses a new series of problems. Once the content is captured as files that can be managed on a computer system the preservation of the files themselves becomes a matter of routine mass storage management in common with all other data held on computer systems. However software is needed to read these files and to present or 'render' the digital data on a computer screen or other viewing device. The software and computer equipment needed to run the software becomes obsolete over time and the ability to read older files may be lost. Special arrangements have to be made to maintain access to the digital content or digital files have to be put into new forms that can be read by current software and displayed on current equipment. Obsolescence is therefore a major problem and digitisation as such does not solve the problems of preservation.

Because the digital files of moving images tend to be rather larger than the files of text documents and other digital objects and hence more time consuming to transfer over the Internet, it is common to present moving image files in a more convenient form, principally through the use of lossy compression techniques. There is a cost to this, since although access is undoubtedly made easier, lossy compression loses some of the original data. The net result is that the needs of access and preservation diverge, to a greater extent than with other common form of digital object, and files that are made available for easy access are not usually the most appropriate ones for preservation purposes.

The general response to these issues is to plan for the longer term, to decide on preservation formats that are likely to be relatively long lived, but also to plan for changes over time. This can be done either by changing the data or by making arrangements so that new software and equipment can continue to access and display data in the old formats. The former approach involves moving with the prevailing software and equipment developments and ‘migrating’ data successively on to current formats. The second means making a series of arrangements and developing software and equipment so that the older programs can be ‘emulated’ in future environments. As we shall see there are variations on these themes of ‘migration’ and ‘emulation’, but whatever the approach, it has to be recognised that changes are inevitable.

It is in this context that the core issue of this study arises. If the original object necessarily has to be changed in the course of processes designed to ensure its preservation over the longer term, which are the characteristics and features of the original that have to be preserved through these changes? In other words, what is the ‘essence’ of the object that constitutes its unique identity and allows us to say that a new representation is the same as the previous one? Or in the terminology that we use: what are the ‘Significant Properties’ of the object for preservation purposes?

The next question is then, how are these Significant Properties to be preserved? It turns out that there are a number of responses to this. At the most basic level, we believe that there is a case for preserving the original data in the original format. We do not know what will be possible in the future.

However, we have just outlined the reasons why simply preserving the original data is not sufficient currently. Some form of migration or provision for emulation is necessary to ensure that the original data can continue to be accessed. If migration is chosen as the preservation strategy, there is a need to check how far the new preservation format captures the Significant Properties of the original and if this is not complete how information about the missing properties can be conveyed. Similarly, if an emulation strategy is chosen, there is a need to check that the emulation software is successfully reproducing the Significant Properties of the original.

Choosing the right preservation format or developing the right emulation software is not the complete answer, then, except in unrealistically ideal conditions. There is a need for another level at which Significant Properties can be preserved. This is essentially the metadata level. Metadata are data about data. Metadata schemes are structures or systems for organising information about data. The question then becomes which metadata schemes have elements that allow information about Significant Properties to be recorded in an appropriate way.

This type of question starts to move the discussion from technical issues to something that is much broader. The instincts of those engaged in preservation are to keep the object as it was originally intended to be. This brings the ‘intent’ of the original creator into the picture as well as the ‘content’ of the object under consideration. More generally, the context of the creation of the object is part of determining what it is, along with the circumstances of its subsequent history. The extent of the considerations to be taken into account in the metadata of a digital object rapidly becomes rather large and a means of distinguishing between them is important for practical reasons.

The dilemma of digital preservation, then, arises from the need to change the original object in order to preserve it, or rather the need to change it successively as the general technical environment continues to change. It is in this context that we have to address the questions of what to preserve and how to preserve it - the core questions we shall address in the following discussion.

2 A Framework for Digital Preservation

2.1 Digital Objects and their Presentation

Before progressing to look at the general framework for digital archiving that provides the basis for much of the work that is being undertaken on digital preservation, it will useful to follow up the discussion in the previous section and set out some fundamental distinctions relating to digital objects and the way that they are used. We begin with digital objects at their basic level:

Digital objects are fundamentally *binary data, encoding* the item of interest as a sequence of binary digits. They may be transmitted, for instance over a network, as a *bit-stream* or stored in a computer system as a *file*. It will be seen that a file containing the digital object is the most basic level at which a digital object can be preserved.

Once stored as a digital file, the digital object can be copied indefinitely without loss, a first step towards digital preservation. The risk of deterioration or loss of the medium in which the file is stored is mitigated by maintaining multiple copies and regular checking, the discipline of *mass storage* management.

However, in contrast to a physical object in an archive, library or museum, a digital object is not accessible in itself. It requires a *player* to interpret or *decode* the binary data and *render* it, transforming it into a form accessible to the viewer presented on an output device, such as a computer screen or loudspeaker.

The file in which the information has been encoded has to be in a recognizable *format* in order for it to be presented successfully by the player. The structure of this file is determined by the format. The file may be a *wrapper* or *package* format, encapsulating multiple bit-steams together with control information and other *metadata*.

There are a wide variety of file formats used to encode digital moving images, many of them proprietary. Different formats and encodings have different properties and are used for different purposes.

In the case of digital moving images we have been through a period of rapid technological change. Broadcast organisations are moving to completely digital systems, consumer camcorders are now largely digital and video streams are seen everywhere on the Internet. This has given rise to rapid developments and consequently rapid obsolescence.

For preservation purposes we want to use formats that can be relied on to be usable in the future. Unfortunately this is not the case, some formats continue and others cease to be supported. The player software may be preserved but the computer hardware or operating environment needed to run the software may not. We cannot assume that any given format will be usable in the future. In fact, on the contrary, we have to assume that all these elements in the processes that allow digital files to be rendered will change significantly over time.

The question therefore is how do organisations concerned with preservation deal with the issue of changing formats, operating software, hardware etc? A list of possible approaches has been provided by Ken Thibodeau of the US National Archives and Records Administration [4]. It includes:

> *Bitstream preservation*: preserving the original data as a sequence of binary digits, either in the original file format or as a sequence of bits contained inside a wrapper. The original data are uninterpreted.
>
> *Technology preservation :* preserving and maintaining the original hardware and software - in effect creating an information technology museum.
>
> *Migration:* moving digital objects onto new media and/or into new file formats. The new file formats can either be 'up-dates of earlier formats or different formats. When the migration of file formats is to a standardised format chosen to ensure future accessibility, the process is referred to as 'normalisation'. Migration risks losing some of the attributes of the earlier format.
>
> *Emulation:* replicating the original experience, including the original 'look and feel', on modern hardware and operating systems. Supporters of emulation believe that preserving the presentation of information is important as well as preserving the content. In this sense, the process is the purest form of archival preservation, but it is very demanding. It requires the development of special software, not only for each format to be preserved, but also for each operating environment in which the objects can be accessed.
>
> *Migration on Request:* preserving digital objects in their original format, while developing migration tools to render the original objects into a current format, when this is required. With this approach, a whole series of migration tools (one for each original format) have to be maintained to keep up with changing computing environments.
>
> *Universal Virtual Computer :* a largely theoretical approach, developed by Raymond Lorie at IBM Research[5] that combines emulation with migration and claims to derive most of the benefits of each approach, while avoiding the major difficulties. As the name suggests, a virtual computer is created, which lies between real computer platforms and the software layers. A UVC thus offers a platform independent layer that can persist indefinitely into the future. In application of the idea, the original object is preserved and it can be reconstructed at any time as it originally appeared.

In practice, the main contenders for attention have been the emulation and migration strategies, We shall see that in the present Study, the migration route is favoured, with bit-stream preservation as an additional back-up.

However, there are still proponents of an emulation strategy. It figured in earlier JISC projects and at a European level, emulation is still very much associated

---

4 Thibodeau, K. 2002. 'Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years' in proceedings of *The State of Digital Preservation: An International Perspective.* Conference Proceedings. Washington. 2002.

5 http://domino.research.ibm.com/comm/pr.nsf/pages/bio.lorie.html

with the Koninklijke Bibliotheek and the Nationaal Archief in the Netherlands and their development of the Dioscuri emulator[6]. These archives are active in the PLANETS project, to which there is further reference below and the issue will be picked up again in the discussion on the section on *Content & Intent*.

2.2 A General Framework for Digital Preservation - the OAIS model

The Open Archival Information System (OAIS) Reference Model provides a general framework for digital preservation. This framework was originally drawn up by the Consultative Committee for Space Data Systems, but it is widely recognised internationally as the basis for digital archiving. It was accepted as an ISO standard in 2003[7].

Archives conforming to the model accept the responsibility to preserve information over the 'long term' and to make it available for a 'Designated Community'. It is understood that the commitment has to be to carrying the information in an accessible form over a sufficiently long period to encounter changing technologies, including new media and data formats, and also to addressing the changing requirements of the user community. The other basic element in the objectives of the archive, as envisaged in the reference model, is that the proposition that the archive needs to specify the type and nature of the community it is to serve – the 'Designated Community' and ensure that archived material is accessible to that community without undue difficulty, particularly in terms of being able to understand the information that is being held by the archive.

One of the greatest challenges initially for archives is finding a shared vocabulary for a common understanding of processes and procedures by the range of stakeholders with differing relationships to the retained material. Indeed one of the objectives in developing the Reference Model was to facilitate the effective participation of non-archival organisations in the preservation process. Considerable effort was therefore devoted to establishing common terms and concepts and providing a framework for setting out the roles of significant players and the relationships between them.

This approach helped to clarify the nature of the processes involved in archiving digital material[8] and, together with the long term view inherent in the model, led to the development of a life cycle perspective on preservation processes, which has been considerably elaborated subsequently. Essentially the life of a digital object is conceived as beginning with the intentions and

---

6 http://dioscuri.sourceforge.net/

7 ISO 14721:2003 Space data and information transfer systems – Open archival information system – Reference model : : http://public.ccsds.org/publications/archive/650x0b1.pdf

8 Note that the OAIS model does envisage that material held in the archive may be in non-digital forms. However, the main focus is on digital information, both as the primary form of information held and as supporting information for both digital and non-digital materials and the modeling and preservation of primary information in non-digital form is not addressed in any detail.

actions of its producer, continuing with modifications made to it by various users, being subject to further modifications in the course of its ingestion into, and management within, an archive and then still further modifications in the subsequent use of the digital object by a 'consumer'. The final stage in the life cycle of the object comes with the decision – conscious or by default – that the object is no longer required and can be destroyed.

There are a number of studies that explore this life cycle theme in considerable depth, including the Digital Moving Images and Sound Archiving Study and other AHDS work[9]. The fact that moving images in particular can easily go through many transformations, often starting in non-digital forms and once there frequently being modified and migrated several times, suggests the particular appropriateness of a life-cycle approach to the understanding of the preservation issues facing archives of moving images and consequently the life cycle perspective very much informs what follows. However, we should be aware that within this overall conception, there will be a particular focus on the archiving process that will leave a number of the wider considerations of the life-cycle model unarticulated in the detail we examine.

The Archival Information Package

The Reference Model provides both a functional model for conducting and managing digital preservation and an information model.

The functional model has been developed further by a number of organisations. Of particular interest is the model for a 'Trusted Digital Repository', developed by RLG and OCLC (Online Computer Library Center)[10], through an international working group. This model sets out the organisational criteria for a responsible digital archive, its attributes and responsibilities, its functions, processes and procedures.

In setting out further the nature of the particular focus of this study, our attention is more on the information model presented by OAIS than the functional model and here it is useful to be able to pick up on some of the terminology and clarifications that were introduced by OAIS.

The information model defines the broad types of information required to preserve and access the digital objects stored in a repository. In particular, It distinguishes between:

- the Submission Information Package (SIP);
- the Archival Information Package (AIP);
- the Dissemination Information Package (DIP)

---

9 Including S Anderson et al 'Digital Images Archiving Study' 2006

10 RLG-OCOC http://www.oclc.org/programs/ourwork/past/trustedrep/default.htm Trusted Digital Repositories : Attributes and Responsibilities 2002

In explaining these it will be useful to reproduce some of the core definitions of the Reference Model :

Terms Defined in the OAIS Reference Model[11]
Information: Any type of knowledge that can be exchanged. In an exchange, it is represented by data. An example is a string of bits (the data) accompanied by a description of how to interpret a string of bits as numbers representing temperature observations measured in degrees Celsius (the representation information).
Submission Information Package (SIP): An *Information* Package that is delivered by the Producer to the OAIS for use in the construction of one or more *AIP*s.
Archival Information Package (AIP): An *Information* Package, consisting of the *Content Information* and the associated *Preservation Description Information (PDI)*, which is preserved within an OAIS.
Dissemination Information Package (DIP): The *Information* Package, derived from one or more AIPs, received by the Consumer in response to a request to the OAIS.
Content Information: The set of information that is the original target of preservation. It is an Information Object comprised of its *Content Data Object* and its *Representation Information*. An example of Content Information could be a single table of numbers representing, and understandable as, temperatures, but excluding the documentation that would explain its history and origin, how it relates to other observations, etc.
Data Object: Either a Physical Object or a *Digital Object*.
Digital Object: An object composed of a set of bit sequences.
Representation Information: The information that maps a *Data Object* into more meaningful concepts. An example is the ASCII definition that describes how a sequence of bits (i.e., a Data Object) is mapped into a symbol.
Preservation Description Information (PDI): The information which is necessary for adequate preservation of the *Content Information* and which can be categorized as *Provenance, Reference, Fixity, and Context information.*
Provenance Information: The information that documents the history of the *Content Information*. This information tells the origin or source of the *Content Information*, any changes that may have taken place since it was originated, and who has had custody of it since it was originated. Examples of Provenance Information are the principal investigator who

[11] http://public.ccsds.org/publications/archive/650x0b1.pdf

recorded the data, and the information concerning its storage, handling, and migration.

Reference Information: The information that identifies, and if necessary describes, one or more mechanisms used to provide assigned identifiers for the Content Information. It also provides identifiers that allow outside systems to refer, unambiguously, to a particular Content Information. An example of Reference Information is an ISBN.

Fixity Information: The information which documents the authentication mechanisms and provides authentication keys to ensure that the Content Information object has not been altered in an undocumented manner. An example is a Cyclical Redundancy Check (CRC) code for a file.

Context Information: The information that documents the relationships of the *ContentInformation* to its environment. This includes why the *Content Information* was created and how it relates to other Content Information objects.

In rather more direct terms, we can say that the *Submission Information Package* (SIP) describes the information, including the relevant discovery metadata, that an archive will request a producer to submit when proposing a data object for preservation (and subsequent dissemination).

The *Archival Information Package* (AIP) is a derived from the SIP and will consist of information that has been manipulated to make it suitable for preservation, especially over the longer term. In an AIP, we would expect to find, according to the OAIS model :

- the digital object (a set of bit sequences);
- information about how to make the digital object meaningful to users;
- a further set of information that is necessary for adequate preservation of the target information, primarily relating to Provenance, Reference, Fixity, and Context information

The *Dissemination Information Package* (DIP) makes use of what is available in the SIP and AIP, but adapts the material in order to deliver it in a form that meets the requirements of a specific group of users.

Our interest includes the nature of the elements of the package delivered by the producer and we have to keep an eye on the requirements of users of material held in the archive, but our focus is primarily on the Archival Information Package and our main task in this study is to determine significant parts of the information set that is necessary in order to meet the second requirement of the AIP and to continue to meet this over time.

We have already seen, however, that there are elements of the third requirement, particularly contextual information that impinge on the task of making the data meaningful, so we shall also be proposing an approach to the coverage of these issues. In order to provide a framework for this, we should note at this point that various approaches have been suggested to elaborating the main elements of the required Preservation Description Information. Most consist of proposing the development of a metadata ‘wrapper’ around the core object data, and specifying the structure and content of these metadata. Unsurprisingly most also propose making use of existing metadata standards and structures.

2.3 PREMIS and Terminology

Before moving on to look at metadata, it is worth mentioning that one of the achievements of OAIS was to help clarify terminology and usage in the preservation area. This process has been taken further in the work of the PREMIS (PREservation Metadata: Implementation Strategies) Working Group[12].

PREMIS was an international group of experts in digital preservation that was set up to develop a core preservation metadata set with broad applicability across the digital preservation community. One of its main achievements was the development of a Data Dictionary[13] that has helped to provide an authoritative definition of much of the core cross-format technical terminology used in the preservation context.

2.4 Metadata wrappers – MPEG-21 and METS

2.4.1 Metadata

Metadata comments on, or provides information about data - in our case, the core digital object - and various metadata schemata provide structures for providing this information in a systematic way. In our context :

> “Metadata accompanies and makes reference to each digital object and provides associated descriptive, structural, administrative, rights management, and other kinds of information.”[14]

Metadata relating to digital objects is more extensive and of a significantly different kind from the metadata used for managing collections of printed works or other physical objects. To provide an orientation in this area, we want to refer to two standards that have been cited in the context of the preservation of moving images - MPEG-21 and METS

Both of these standards have been highlighted by S Wilson et al in *Digital Moving Images and Sound Archiving Study* as suitable ‘container schemas’ The Study comments that ‘METS is emerging as the metadata of choice for many digital repositories’, while ‘specialist audio-visual archives *may* be more disposed to MPEG-21’. It then adds that ‘more general (e.g. University) repositories that archive limited amounts of audio-visual data might be expected to gravitate more towards METS because of its library origins’.

---

12 http://www.oclc.org/research/projects/pmwg/

13 http://www.oclc.org/research/projects/pmwg/premis-dd.pdf

14 Clifford Lynch (D-Lib Magazine, 1999)

Section 7.5 of the Study does consider a number of other metadata standards relevant to moving image (and audio) preservation, along with further detail on MPEG-21 and METS.

2.4.2 MPEG-21

The MPEG-21 standard was developed by from the Moving Picture Experts Group, the ISO/IEC working group, that is responsible for developing international standards in the video and audio area. It is a machine-readable, XML-based, standard that allows license information to be accessed in an 'ubiquitous, unambiguous and secure' manner.

The aim with the development of MPEG-21 was to contribute to the control of the the distribution of multimedia digital content and especially to help stop illicit file sharing. It is therefore concerned to a large extent with expressing, organising and distributing information about intellectual property rights – an issue of some complexity in the case of moving images.

Although this is undoubtedly an issue that continues to assume an ever-increasing profile, the management of intellectual property rights does not in itself form part of the target area of our particular study. Nonetheless, MPEG-21 is of interest, since in defining a framework for distribution of digital objects, such as digital moving images, the standard addresses the needs of a wide range of actors, from original content creators and producers to distributors and various other 'players in the delivery and consumption chain'. In doing so, it requires a package to brought together that includes many of the elements that are of interest in the context of the preservation of significant properties.

A central concept in MPEG-21 is that of a 'digital item'. Digital items are structured digital objects that combine the core components of the object – individually identifiable data streams, known as 'resources', such as video sequences, audio tracks and separate images - together with a structure that defines the relationship between the resources and fairly extensive metadata. The framework can be expressed at various degrees of granularity and allows for considerable flexibility, including the incorporation of pre-existing metadata, held, for instance in a METS format. MPEG-21 is thus a wrapper that ties together the disparate elements needed for providing widespread access, but also many of the elements required for preservation.

The main problem with MPEG-21 is the apparent complexity, which is the other side of its flexible and comprehensive structure. It is for this reason that its use is likely to continue to be restricted to specialist audio-visual archives or organisations that are actively engaged in distributing archived material to a range of diverse users and consequently need to make use of the rights management features of MPEG-21.

2.4.3 METS

METS - the Metadata Encoding and Transmission Standard – takes us back to the slightly broader context of digital repositories. It is a standard that was developed principally by major library and archiving institutions in the United States to provide a consistent approach to formulating descriptive, administrative, and structural metadata describing digital objects within a digital library environment. As such it relates to the range of digital objects held by such organisations, being specifically designed to be applicable to audio and video files as well as to electronic texts and still images.

METS supports both the management of digital objects within the repository and their exchange and use externally. It structures the metadata in a form that is intended to facilitate sharing, cross-searching and exchange and to allow digital objects to be easily rendered for browsing and display purposes.

METS is expressed in XML, which means that it conforms to a widely recognized standard, designed to facilitate access to documents, their transfer and their automatic processing, by means of a standard for labelling the components of these documents in a systematic way, thus enabling these components to be recognised in terms of their function within the document.

The METS Reference Manual[15] describes seven major sections in the METS schema :

The major sections of the METS schema
METS Header – The METS Header contains metadata describing the METS document itself, including such information as creator, editor, etc.
Descriptive Metadata Section – This section contains descriptive metadata that is external to the METS document (e.g., a MARC record in an OPAC), internally embedded descriptive metadata, or both.
Administrative Metadata Section – Information about how the files were created and stored, intellectual property rights, metadata regarding the original source object from which the digital object was derived, information regarding the provenance of the files that comprise the object. As with descriptive metadata, the administrative metadata can be either external to the METS document, or encoded internally.
File Section – A list of all files that contain content which make up the electronic versions of the digital object.
Structural Map – This is the heart of the METS document. It outlines a hierarchical structure for the digital object, and links the elements of that structure to content files and metadata that pertain to each element. The structural map is the one mandatory section in a METS document.
Structural Links – Allows the creator of the METS document to record the existence of hyperlinks between nodes in the hierarchy outlined in the Structural Map.
Behaviour Section – A behaviour section can be used to associate executable behaviours with the content of the object encoded using METS.

15 <METS> Metadata Encoding and Transmission Standard: Primer and Reference Manual
http://www.loc.gov/standards/mets/METS%20Documentation%20final%20070930%20msw.pdf

Within this framework however, there is considerable flexibility with respect to the detail of the metadata to be provided. The interest of this for us is that it can be made to conform to the requirements of the OAIS model and provide the (mainly common) elements needed for Submission Information Packages, Archival Information Packages and Dissemination Information Packages. In the case of AIPs, for instance, it can provide a structure for the metadata on Provenance, Reference, Fixity, and Context information that make up the necessary Preservation Description Information.

An interesting exercise in assessing the applicability of both the OAIS Reference Model and, within this, the METS schema in a UK context, was undertaken in a JISC-funded study[16] by The National Archives (TNA) and the UK Data Archive (UKDA) at the University of Essex in 2004. This 'Assessment of UKDA and TNA Compliance with OAIS and METS Standards' necessarily confronted the practical issues that arise in implementing such a model, and, in their case, in judging the compliance of procedures and systems that pre-dated the development and publication of the reference model. It therefore provides a wealth of useful detail on implementation and practical assistance for other institutions[17], as well as much more information on the Reference Model and METS than it has been possible to present here.

The conclusions of the report on the project point to a series of difficulties in implementing the OAIS model, including difficulties in operationalising the concept of a 'Designated Consumer Community', and some departures from implied practice. Both TNA and UKDA are keen, for instance, to maintain the archival tradition in which original records are often copied and only the copies are made available to users, whereas the model seems to assume that DIPs are created 'on the fly' from the AIP in storage. Of most relevance for our purposes, however, is the observation :

> 'In some cases, it has been possible to see where the OAIS model does not provide enough detail to make good statements about the archive's operation. This is particularly so in the areas of ingest, access and migration. This is not a weakness of OAIS as it is attempting to be very neutral about the actual preservation and access implementation methods but it does mean that OAIS does not provide a conceptual map or a common terminology for some very important digital preservation areas common to both TNA and UKDA, and to many other archives.'

It is in this area that the current study is attempting to elaborate on earlier work.

One final comment on the UKDA/TNA project comes from their remark that one of the unexpected outcomes of the work was the utility of the OAIS language as a means of communication between the staff of the institutions involved. We feel that this reinforces our belief that continuing to set out and clarify a conceptual framework could be as important as arriving at the right technical conclusions in a study of this kind.

---

16 'Assessment of UKDA and TNA Compliance with OAIS and METS Standards': http://www.data-archive.ac.uk/news/publications/oaismets.pdf

17 It includes, for instance, a set of questions for self-testing for OAIS standard compliance

2.5 Project CASPAR

Some of the most recent developments in this are taking place within a framework established under CASPAR, a European Union research project. CASPAR – 'Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval' - is an Integrated Project that started on 1 April 2006 and has, like other projects of this kind and scale, an objective of bringing about much greater co-operation and co-ordination of research activity in this area at a European and indeed a global level.

The aim of CASPAR is to 'research, implement, and disseminate innovative solutions for digital preservation based on the OAIS reference model'. The core framework of OAIS, and in particular the Functional Model and the Information Model, is accepted, but the intention is to supplement and extent this framework, especially in areas that are not covered in depth by OAIS, for example, by addressing issues of preservation over time, including Digital Rights Management and issues of Authenticity. CASAR is moving in the direction of the Significant Properties studies, within the OAIS framework, and in addition, it is looking at broader issues of digital asset management and also addressing issues of implementation by testing the OAIS model against real case studies, elaborating metadata frameworks and developing tools and techniques to support the capture and use of the relevant information.

Progress with this project is still difficult to judge at this stage, but if it succeeds it will contribute a major new element to the framework and day-to-day context, within which the determination of Significant Properties has to take place.

Interested parties are able to register their interest on the CASPAR web site: http://www.casparpreserves.eu/publications/deliverables

2.6 The Requirements of Digital Preservation

It will be useful, we hope, before going on to look more specifically at the characterisation of significant properties to provide a series of summary statements that set out the main elements derived from our consideration of archiving models that we will need to carry over into the rest of the study. These are:

- The preservation of digital objects, in the present case, digital moving images, takes place within the broader context of the processes and procedures of digital archiving, for which the Open Archival Information System Reference Model provides an internationally recognized framework and vocabulary.
- Within the OAIS model our main focus is on The Archival Information Package (AIP) which consist of :
    - the digital object (a set of bit sequences);
    - information about how to make the digital object meaningful to users;

    - a further set of information that is necessary for adequate preservation of the target information, primarily relating to Provenance, Reference, Fixity, and Context information
- We can look to the METS metadata schema to provide us with a framework that is compatible with OAIS and that, at the same time, allows us to associate with the digital object information on some of its broader properties.

3 The Framework for Determining Significant Properties

3.1 Context – InSPECT, CEDARS and CAMiLEON

3.1.1 Project InSPECT

The immediate framework within which we are setting out the significant properties of digital moving images is the JISC-funded project: Investigating the Significant Properties of Electronic Content Over Time (InSPECT)[18]. This is being undertaken by the Arts and Humanities Data Service (AHDS), in association with The National Archives (TNA) and aims to articulate the significant properties of a series of digital objects (raster images, emails, structured text, digital audio) but also to develop the methodology and provide a basis for the further projects that are extending the range of digital objects under consideration by examining the significant properties of e-learning objects, software, vector images and in our own case, moving images. InSPECT therefore provides a framework, within which the other projects have been encouraged to work and, certainly, one of our aims is to develop an approach that is consistent with that adopted by InSPECT.

InSPECT builds on two earlier projects funded by JISC: CEDARS and CAMiLEON

3.1.2 Project CEDARS

CEDARS[19] has been a highly influential project in developing ideas on digital preservation and much of what it had to say has been reflected in the previous sections. Starting from the OAIS model, it developed the idea of Content Information (referred to as 'resources') being packaged together with metadata – Preservation Description Information (PDI). CEDARS then made use explicitly of the concept of Significant Properties in terms of bringing together the Representation Information relating to a digital object and the associated PDI – 'all of the other supporting metadata that is deemed necessary for purposes of long-term preservation'. The Representation Information relates, (in the OAIS model, it will be recalled), to the key element in digital preservation – the information that supports the transformation of the Primary Digital Object into a usable digital object. CEDARS therefore settled much of the basic framework for discussing Significant Properties.

CEDARS also articulated in clear form the fairly fundamental notion in digital preservation that is carried through into the InSPECT project, namely that it is never possible to entirely reproduce all aspects of the presentation of the original object. A degree of choice is inevitable with respect to those aspects

---

18 http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/inspect.aspx

19 The Cedars Project Report (April 1998–March 2001), June 2001, p 14 http://www.leeds.ac.uk/cedars/admin/CedarsProjectReportToMar01.pdf

that are a matter of priority in the preservation process. There has to be a choice about which properties are significant and this is precisely the task required in the characterisation of Significant Properties.

Further information on the CEDARS project can be found at:

http://www.leeds.ac.uk/cedars/

3.1.3 Project CAMiLEON

The CAMiLEON project further developed a model for expressing Significant Properties, again within the OAIS framework. Central to the study was a review of properties already highlighted in standards and frameworks established by bodies such as MPEG and NISO as a basis for identifying Significant Properties. Some 800 property references were identified, extending to stylistic and aesthetic features. In view of this wide range, CAMiLEON refrained from recommending a single set of decision-making rules for determining Significant Properties, but the process was seen to be helpful in articulating the available options and in documenting the trade-offs among them.

Two particular features of the CAMiLEON project are of particular interest for the present study :

- First, there is the fact that CAMiLEON explored the extent to which properties could be regarded as significant, giving consideration to stylistic and aesthetic features and concluding that their significance was very much tied to the context of their creation and use.
- The second was that the CAMiLEON project aimed to assess emulation as a preservation strategy and devoted a fair amount of its resources to addressing this question. It concluded eventually that although emulation is technically feasible, it is not always the most effective or cost-effective strategy.

Further information on the project, including project reports, can be found at:

http://www.si.umich.edu/CAMILEON/

A brief overview[20] of some of the issues is to be found at:

http://www.ils.unc.edu/callee/sigprops_dlm2002.pdf

3.2 The Central Features of InSPECT

3.2.1 The Performance Model

InSPECT builds on these CEDARS and CAMiLEON but also owes a lot in its orientation to work by the National Archives of Australia (NAA) and The National Archives (TNA) in the UK.

---

20 Margaret Hedstrom, Christopher A. Lee, *Significant properties of digital objects: definitions, applications, implications*, 2002.

In particular InSPECT adopts from the NAA a distinctive view of the way that content information interacts with the media that store it and the equipment that displays it. This is known as the ‘Performance Model’.

The Performance Model sees each interaction between a data source (‘the source’) and the technology that presents it (‘the process’) as a ‘performance’, by analogy with the way that a play or a piece of music is performed.

Furthermore it recognises that each combination of source and specific process platform may produce a slightly different performance from other combinations. Change is seen as inevitable. The need in a changing environment is to determine what is essential and what is not. Citing earlier work by TNA, InSPECT asserts that

> ‘*a record is considered to be essentially complete and uncorrupted if the message [it] meant to communicate in order to achieve its purpose is unaltered.’*

This puts the definition of Significant Properties in the centre of the picture, because it is the role of Significant Property analysis to define those properties that it is necessary to reproduce in order to provide an ‘authentic’ performance – a performance that is essentially the same as the original. A key challenge in digital preservation, then, is precisely to determine Significant Properties.

Core documents describing the InSPECT approach are A Wilson *Significant Properties Report*. 2007[21] and G Knight, *Framework for the definition of significant properties*, 2008.

3.2.2 Migration

Accompanying the ‘performance’ conception is the view that the role of Significant Property analysis is to define those properties that it is necessary to reproduce in order to provide an ‘authentic’ performance. It is understood that with this process, digital objects cannot be fully preserved in their original state. Indeed, InSPECT argues that this is not necessary:

> *‘neither the source nor the process need be retained in their original state for a future performance to be considered authentic. As long as the essential parts of the performance can be replicated over time, the source and process can be replaced’*

The issue again becomes one of defining the characteristics that need to be reproduced in order to preserve the authenticity of the original object.

This opens the way for a ‘normalisation’ strategy. The idea is to select for each type of digital object a very small range at any one time of open, non-proprietary digital formats that have the prospect of reasonable longevity and have well-known properties. A central part of the preservation process then involves migrating digital objects onto this standard format (or formats) that then becomes the core of the Archival Information Package, in OAIS terms. Over time, as technology continues to change, there will be further migrations onto new standard formats.

---

[21] http://www.significantproperties.org.uk/documents/wp22_significant_properties.pdf

The role of an established set of Significant Properties in this context, is to act as a check-list, by which it can be ascertained that all the important characteristics of a digital object have been migrated in the course of the move from one format to another.

InSPECT suggests that migration is the sole strategy to be adopted. This is a coherent position, but actually it is a view that we would wish to modify to the extent that we believe that there is some merit in retaining, where possible, the original data in its original format. This might be a matter of 'belt and braces', but it is also possible to conceive of circumstances in which the original data could be of interest in the future.

3.3 Development of the InSPECT Model

The work of InSPECT is continuing, as this report is being written, and there have been a number of developments[22], building on the framework already outlined that we have to take into account, if we are to have a consistent approach applied across all the Significant Properties projects.

In particular, while continuing to make use of the basic concepts already outlined, in carrying out detailed work on particular digital objects, InSPECT is bringing in concepts and processes developed in other major projects, including:

- the Seamless Flow programme[23], which is providing a framework at The National Archives for the digital preservation of public documents and other material from across UK government departments;
- the PREMIS (PREservation Metadata: Implementation Strategies) Working Group[24], an international group of experts in digital preservation that was set up to develop a core preservation metadata set with broad applicability across the digital preservation community;
- the Digital Preservation Testbed project that helped establish the preservation strategy of the Nationaal Archief of the Netherlands.

Part of the aim is to arrive at practical processes for the determination of Significant Properties that make use, as far as possible, of existing metadata. A simplified framework is proposed providing a common set of metadata elements similar to the Dublin Core metadata standard. It is envisaged that the metadata set could be used by an institution to indicate the properties that are considered to be important by that institution and its Designated Community and to indicate the quality thresholds that must be met.

This simplified framework for defining significant properties is said to have two benefits:

---

22 Described in Gareth Knight, Framework for the Definition of Significant Properties, AHDS 2008.

23 http://www.nationalarchives.gov.uk/electronicrecords/seamless_flow/programme.htm

24 http://www.oclc.org/research/projects/pmwg/

- Documentation on the significant properties of a Record may be stored with the digital Record itself and transferred between different digital repositories;
- The significant properties may be used to quantitatively measure automated conversion processes.

In order to arrive at this, the InSPECT team again reviewed format specifications and their key components (where objects can have more than one constituent component - text, images, sound etc.), relating to the four object types, being considered directly by the project - raster images, digital audio, structured text and emails. As a result of this exercise, the team concluded that it is useful to refer to the following tasks in progressing with the analysis ;

- Definition of the *intellectual components* that the assessor believes should be maintained;
- Identification of the *technical properties* of each component that are required to recreate the component(s);
- *Classification of the function* performed by each property and an assessment of its value;
- *Measurement of the technical and intellectual properties of the digital Record*, as well as consideration of the requirements of the institution and Designated Community.

In referring to the 'intellectual components', the team have in mind the broader context for understanding what an object is or does and the relationship between components, where the object has more than one element – images and sound track, for example. Defining intellectual components involves taking into account the standards with which the institution concerned is required to comply and the requirements of the Designated Community, but overall, it involves a decision about 'the component's contribution to the recreation of the record as a 'whole'.

With regard to technical properties, there are three well-established principles that have to be applied :

- *Function*: what is the function that the property performs in relation to the Component or the object in its entirety?
- *Uniqueness*: Do other properties exist that perform a similar function and could be substituted?
- *Robustness*: What effect will it have on the recreation of the object if the property is damaged?

With the measurement of properties there are two issues that have to be taken into account. Specific properties of moving images can have different values and these values have to be established and recorded. That is clear and straightforward.

There is a second issue, however, to do with the measurement of the relative significance of properties. For other digital objects, there can be varying degrees of significance in different applications. In the area of vector graphics, for instance, there are a large number of potential properties and these can often have differing significance depending on the particular application. The colour of a line may or may not be significant. Recognizing these degrees of relative significance, a parallel study on the Significant Properties of Vector Images, conducted by System Simulation and colleagues at Oxford Brookes University[25] proposed a metric consisting of a 5-point scale. The InSPECT project adopts a similar approach.

In the case of moving images of the type under consideration (this may well be different for medical or scientific moving images), this type of variation in significance does not arise to anything like the same extent. The rendering of moving images – their projection – does not allow their presentation to vary anything like as much as is possible with, say, the rendering of an e-mail message or a vector graphic. The apparent exception to this is aspect ratio, but even here we are not concerned with degrees of significance. The aspect ratio intended by the producer is the Significant Property. This has to be reproduced if the authenticity of the copy is to be maintained. Subsequent modifications of aspect ratio are derived versions that may or may not be of interest in their own right.

We would conclude that, while recognizing that measurement of (varying degrees of) significance is part of the overall framework defined for our project by InSPECT, it is an element that will have a dormant status in the application of the framework in the area of moving images.

3.4 The Taxonomy of Properties

Bringing together the elements of the model that has been elaborated, the InSPECT project team drew up a taxonomy of properties that are intended to serve as the basis for an exercise in analysing Significant Properties. These are defined as follows:

- *Content:* an abstract term to describe the expression of intellectual work. In a digital environment, Content may describe text, still and moving images, audio, and other intellectual productions. Examples (digital audio): *Duration, Number of channels.*
- *Context:* information that describes the environment in which the Content was created or that affect its intended meaning. Examples (all): *Creator, Creation date, Title*
- *Rendering*: information that contributes to the recreation of the message. It may be applied to a visual or audible component that is a part of the message. Examples (digital moving images): *Resolution, pixel aspect ratio, frame rate*
- *Structure:* information that describes the relationship between two or more types of Content, as required to reconstruct the performance. It

---

[25] M Coyne et al, The Significant Properties of Vector Images, 2007.

may be applied to the intrinsic or extrinsic relationships contained in the message. Examples (digital audio): *Channel relationship*

- *Behaviour:* information that describes the method in which the Content interacts with other stimuli. Stimuli may include the interaction of the user with the software, or interaction with other sources of information, such as an external resource, that affects the context, content, structure, or appearance of the resource. Behaviour is considered to be the most difficult characteristic to preserve – it is often tied to the capabilities of particular software applications and may be difficult to translate. It is also difficult to define all behavioural characteristics in a quantitative manner. Examples (digital documents)*: hyperlinks.*

We can see here a further development of the OAIS framework that not only provides more structure for significant preservation information, but also takes us further in the direction of practical implementation. This move reflects another element in the latest framework developed by InSPECT, in that a data model is proposed, to be used in capturing the necessary information. This data model draws on work in the PREMIS project and more particularly on the Seamless Flow programme, but again aims to achieve simplification, where possible, while allowing enough flexibility to allow descriptions of a range of object properties at different levels of granularity.

The abstract data model is presented as Appendix A, and its application to digital moving images in Appendix B.

It is also worth noting at this point that the InSPECT Taxonomy also has a broad correspondence to METS metadata categories. This provides scope for an accommodation between the InSPECT Taxonomy and METS for those institutions that have a policy of using a METS framework.

3.5 PRONOM – A registry of resources

PRONOM is a registry of resources relating to digital preservation developed and hosted by The National Archives (TNA). It is accessed through TNA's web site: http://www.nationalarchives.gov.uk/aboutapps/PRONOM/tools.htm

This already provides various services and tools, including access to DROID (Digital Record Object Identification), which performs automated batch identification of file formats.

The PRONOM site also provides links to similar projects around the world, including that of the Harvard Object Validation Environment, whose JHOVE tool allows the automatic identification, validation and characterisation of a range of digital object types, see http://hul.harvard.edu/jhove/index.html

3.6 A European context – PLANETS

Before finally setting out the specific framework that will be employed in the definition of the Significant Properties of moving images, there is one last consideration that can best be explained in the context of a European project – PLANETS.

PLANETS (Preservation and Long-term Access through Networked Services)[26], is a European Union IST project that aims to build practical services and tools to reduce the costs of digital preservation and to help ensure long-term access to digital cultural and scientific assets. It started in June 2006 and one of its interesting features is that it is bringing together for further development a number of notable elements that already have a track record at a national level. These include the emulation tools developed by the Koninklijke Bibliotheek and Nationaal Archief in the Netherlands and work by The National Archives from the UK.

PLANETS aims to deliver:

- Preservation Planning services that 'empower organisations to define, evaluate, and execute preservation';
- Methodologies, tools and services for the characterisation of digital objects;
- Innovative solutions for Preservation Actions tools which 'will transform and emulate obsolete digital assets'
- An Interoperability Framework to seamlessly integrate tools and services in a distributed service network
- A Testbed to provide a consistent and coherent evidence-base for the objective evaluation of different protocols, tools, services and complete preservation plans
- A comprehensive Dissemination and Takeup programme.

Understanding Significant Properties is said to be an important element within PLANETS and part of the aim of the Preservation Planning sub-project is to develop practical methodologies and tools for evaluating alternative preservation actions against standard criteria, including evaluating the ability of these actions to recognize and reproduce Significant Properties.

The Characterisation sub-project is led by The National Archives (TNA) and will provide a generic framework to deploy new and existing characterisation tools automatically and to process their outputs in a standardised manner. These characterisation tools will be able to be used to measure the Significant Properties of specific digital objects.

In this respect, PLANETS will be developing the tools that are made available by the PRONOM registry.

---

26 http://www.planets-project.eu/

4 Moving Images & Their Digitisation

4.1 Moving Images

Moving images consist broadly of a sequence of still images projected at a rate that the eye and brain interprets as continuous motion. Moving images in their original form are carried by a variety of media. The principal non-digital forms are:

> *Film*: moving images recorded by a film camera, together with associated sound tracks. For over a century, the predominant way of making moving image works was through the exposure of a sequence of frames of celluloid film and, after processing and editing, the storage of the resulting image sequence on a reel, ready for projection.
>
> *Broadcast television*: moving images and sound delivered by means of a broadcast signal, picked up by a receiver. Broadcast television is a real-time transmission and must be recorded to be archived.
>
> *Video*: technology originally developed for recording broadcast television on tape, but has since been used in various formats for a broader range of applications, including direct video recording.

Increasingly the digital counterparts of the original technology are replacing these legacy formats. Consumer video technology is now largely digital, the Library of Congress notes:

> *"This is written at a time of profound change for the industries that produce moving image content. Broadcasters are in the throes of the … switch to Advanced Television Standards Committee (ATSC) digital transmissions, as cable and satellite distribution entities make parallel changes. At the same time, moving image production and post-production (shooting and editing) is moving from film and tape to disk, server, and workstation, while digital formats will be employed for theatrical distribution in the foreseeable future."*[27]

Accordingly Archives will be called upon to handle digital SIPs as well as those in pre-digital formats.

Medical and scientific imaging refers to a series of techniques and processes that produce still and moving images of biological and physical structures and visualisations of processes. The technologies and engineering required to produce and present these images are diverse and complex. In many cases the images may be recorded using film or video or digital video formats. Where they employ proprietary or specialized digital formats they are regarded as beyond the scope of this study.

4.2 A sequence of still images

Conceptually, moving images are recorded as a sequence of still images, or *frames*, each capturing the scene at a point in time. By displaying the sequence

---

[27] http://www.digitalpreservation.gov/formats/content/video_quality.shtml

of images at the appropriate the illusion of a moving image is created. At an abstract level the moving image can be characterised by the properties relating to a single frame together with the rate at which the images are to be displayed, the *frame rate.*

The images may be accompanied by one or more audio channels and supplemented by additional metadata. The whole may be encapsulated in a suitable*wrapper format* and held in a file.

4.3 Application of the Performance Model

To perform a recording of moving images stored in a digital format, the file must be decoded by appropriate software to produce the audio and image bit-streams. The image bit-stream is processed to generate a series of images on a display at the appropriate frame. The audio bit-stream is processed by a digital-to-analogue converter and the resulting signal played back through one or more loudspeakers. Accompanying metadata can also be presented.

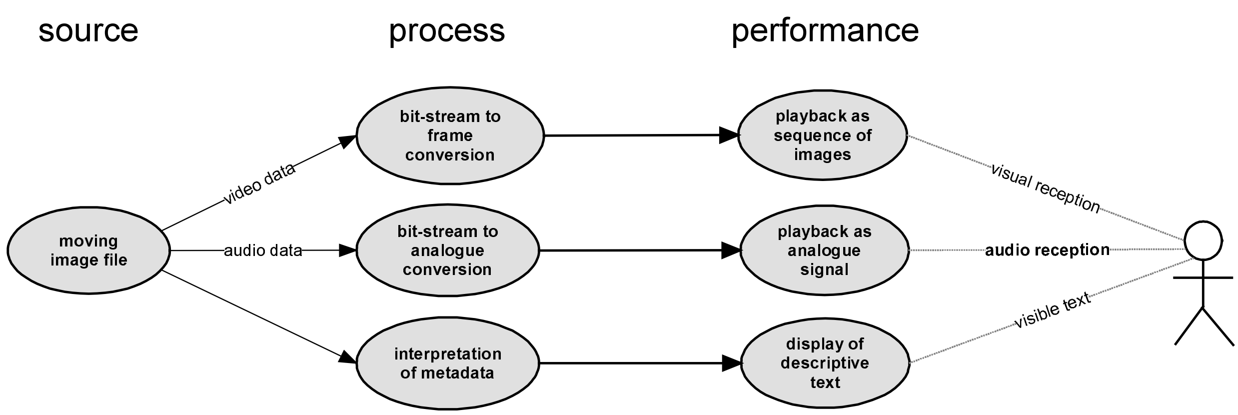


Figure 1. An illustration of the process required to interpret moving image data and render it for the user.

Digital Moving Images and Sound Archiving Study provides a clear and useful overview of the technical and physiological processes behind our 'seeing' the various forms of moving image. This overview is recommended as a fuller statement of the description of the basic processes involved in the creation and rendition of moving images.

4.4 The Significant Properties of Moving Images

4.4.1 Overview

It is appropriate at this stage, to pull the preceding discussion together and move towards a definition of the Significant Properties of moving images following the classification adopted by the InSPECT project. These are the properties that need to be preserved to ensure an 'authentic' performance.

Given that moving images consist of a series of single images, the basic technical features that characterise moving images are:

- The size and shape of the individual images;
- The detail in which the images are recorded;
- The speed at which they follow each other in sequence, when presented;
- Any accompanying audio;
- The overall length of the sequence.

Other properties describe the context in which a digital object has been created and its subsequent history. We shall postpone some of the aspects of information relating to context to the chapter below on *Content & Intent in Moving Images*, but it is worth noting at this point that at a minimum, some basic contextual information is necessary to record the creation and provenance of the work.

4.4.2 Content

Channels: Digital moving images consist of a succession of still images that must be displayed at a given rate to create the illusion of movement. In some cases there may be more than one stream of images (for instance to overlay subtitles or to allow the creation of a 3D illusion with suitable projection). There may also be one or more accompanying audio channels:

Image streams	number of channels	dimensionless
Audio streams	number of channels	dimensionless

Length: The length of the moving image sequence can be measured as the number of frames:

Length	number of frames	dimensionless

The length can also be measured as the duration in seconds. Duration and the number of frames are related by the frame rate:

$$\text{duration} = \text{number of frames} \times \text{frame rate}$$

4.4.3 Context

The file format used to store the digital object may store metadata that can assist with the understanding of the message but are not essential to its access.

In defining the Significant Properties of audio files, the InSPECT project identified *Creation date, Creator* and *Title* as necessary elements. These three properties are taken from the Dublin Core Metadata Element Set (DCMES),[28] a widely used set of descriptive elements for aiding resource discovery.

Others, including the Digital Moving Images and Sound Archiving Study, have proposed using additional elements from the DCMES. The Study, in a slightly

[28] ISO 15836 or ANSI/NISO Z39.85 –2007 : http://www.niso.org/standards/standard_detail.cfm?std_id=725

broader context, suggests *Subject, Identifier* (an unambiguous reference number), *Provenance* and *Rights*.

Provenance is defined as 'A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity and interpretation'. Provenance is of particular interest as knowledge of the nature of the original source material may inform subsequent performances.

While Identification and Rights are undoubtedly important elements in the overall management of preservation, in our view, they are not a core part of the process of reproducing digital material and therefore should not be considered a Significant Property.

The other elements mentioned, however, are part of that process and our conclusion is that a minimal set of context properties consists of: *Title, Creator, Subject, Creation date* and *Provenance.* Going beyond this, DCMES-level of metadata, the choice of the metadata to record will be highly dependent on the nature of the work.

If the archive has the resources to devote to enriching the metadata they can adopt additional elements from the DCMES to describe the work as a whole. The MPEG-7 standard[29] provides a mechanism for describing multimedia content multiplexed with the content itself. Thie enables, for instance, the lyrics of a song to be linked to a video of an artist performing the song or the names of actors and the characters appearing in individual shots in a film to be linked to the performance of the shot itself. The PrestoSpace project Work Area *Metadata and Delivery* [30]developed the use of MPEG-7 in the automated acquisition of metadata relating to recordings of news broadcasts. The OpenDrama project developed an MPEG-7 classification scheme for recordings of operatic productions.[31] The use of a system such as MPEG-7 allows the content to be hugely enriched with contextual information, however the cost of providing this will be significant and, in general, not easily automated.

A measurement of the metadata contained in a digital object will:

- Identify if the digital object contains metadata
- Identify the metadata fields and whether they are populated
- Record information on the content contained in each field

4.4.4 Rendering

Resolution: An individual frame from the digital moving image sequence can be considered as a raster image composed of a rectangular array of *pixels* each representing a discrete point in a two-dimensional space. The spatial *resolution* is determined by the number of pixels making up the frame.

---

29 MPEG-7 Overview, ISO/IEC JTC1/SC29/WG11N6828 http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

30 http://digitalpreservation.ssl.co.uk/metadata/; http://prestospace.org/training/publications.en.html#MAD

31 http://mtg.upf.edu/opendrama/docs/mpeg7/OpenDrama_ClassificationSchemes.html

The range of colours that can be distinguished in the frame is determined by the maximum value that can be held for each pixel, this is referred to as the *bit-depth,* the number of bits used to represent the value of the pixel.

Width	number of pixels	dimensionless
Height	number of pixels	dimensionless
Bit-depth	number of bits	dimensionless

Gamut: The appearance of each pixel, its colour and brightness, is determined by three values, or components, in a *colour model* that map to a colour in a *colour space.* The colour model determines how the values of the components are encoded in the available bit-depth.

Colour model	name	enumerated set
Colour space	name	enumerated set

Pixel aspect ratio: The pixels may not be square, it this case to render the frame correctly the aspect ratio of the pixels must be taken into account.

Pixel aspect ratio	ratio	dimensionless

Frame rate: The rate at which the individual frames are to be displayed.

Frame rate	+ve number	frames/second

Audio aspects: One or more channels of audio material may accompany the sequence of images. The Significant Properties of these channels as digital audio objects are assumed to be the same as any other digital audio material (see G. Knight et al, *Framework for the Definition of Significant Properties*).

4.4.5 Structure

The structure of a moving image sequence with accompanying audio is implicit in the content. The following relationships can be identified:

- The relationship between the audio and video channels must be maintained to ensure that the content can be rendered correctly.
- The relationship between the audio and video streams and the metadata must be maintained to provide the correct context.

4.4.6 Behaviour

No behavioural aspects of digital moving images were identified. Some digital broadcasts and DVD products have interactive features which would exhibit behavioural aspects, however these forms of media are considered outside the scope of this study.

4.5 Influence of the source

4.5.1 From digitisation to preservation

Digitisation is mainly driven by the desire to preserve heritage film and video stock that is perceived to be under threat to a greater or lesser extent and indeed, the process is often accompanied by delicate restoration work. The focus of our interest is somewhat different. We are more concerned with the preservation of the resulting digital stock into the future and ensuring that when digitised material is deposited in archives, it is in a form, and is accompanied by the right information, to allow it to continue to be accessed and used and also to allow it to be migrated to new storage formats, as technology changes. As we have seen, defining the preservation format and the necessary accompanying information is at the heart of what we mean when we refer to the definition of Significant Properties.

Accordingly, the above analysis proceeded from a consideration of what is needed in order to store a sequence of digital moving images for preservation purposes. It deliberately ignored any consideration of the source of the moving images other than information relating to provenance. Clearly, however, the two concerns of digitisation and preservation are very closely related and in practical terms, it comes down to ensuring that future preservation needs are taken into account at the time of digitisation. This means that steps have to be taken to capture all the relevant properties of moving images that will allow us to say that a future viewing of a moving image sequence is essentially the same as the original experience.

The contextual metadata described above allows us to describe the source and its provenance including any digitisation process undergone. This may help recreate more authentic performances. However such metadata is generally for human interpretation. The following sections examine the question of whether there is any information relating to the source that can be stored in a quantitative fashion to improve the authenticity of the performance.

4.5.2 Film

The properties of an image recorded on film are determined by the film stock and camera used as well as the actual exposure process. The **spatial resolution** is determined by the chemistry of the film and the size of the recorded image. A variety of film stocks have been used supporting a range of different frame sizes. These can be digitised frame by frame by a scanning device.

The range of colour and the colour model are all determined by the chemistry of the film. The **gamut** of the film may be greater than that afforded by the digital recording.

The **frame rate** is not recorded in the film itself (although it is implicit in any audio track) but is inherent in the camera and projection system used. It needs to be supplied as part of the digitisation procedure.

The image **aspect ratio** is determined by the film stock, camera and lens. Films in an **anamorphic** format are filmed with a lens which distorts the image aspect ratio to record a wide-screen image onto a standard format film. When digitised, such images can either be recorded as having non-square pixels or can be processed in an analogous fashion to the anamorphic lens to use square pixels.

Some film formats (eg Cinerama and some 3D systems) require the use of more than one projector. These can be recorded as **separate image streams**. They require that attention is given to synchronising these image streams

The **audio information** may be recorded as an optical or magnetic track on the film. This can be played back and digitised in the same manner as other analogue audio. Attention has to be given to ensuring that the audio can be synchronised with the images.

These **Significant Properties** are all preserved in the proposed model within the constraints observed above.

The original performance of the film may have been degraded by **imperfections** in the camera or projector. For instance, a fairly common effect is caused by successive frames not being perfectly aligned in the vertical direction resulting in the image 'jittering' up and down. If the effect was produced by the projector it will not be captured on digitisation. If the effect was produced by the camera, it could be captured in the digitised image itself or by recording metadata describing the misalignment. Whether this is regarded as a Significant Property or not will depend on the policy of the archive with regard to restoration.

Once digitised other imperfections such as scratches or dirt can be removed automatically. However these will result in some other degradation of the image so the policy may be to keep the original digitised images as well as the restored versions. Metadata about detected faults and **restoration** can be recorded with the images. The PrestoSpace Work Area *Restoration*[32] addresses these issues.

4.5.3 Video

Video recording was developed as a way of capturing a television signal, itself an **elusive phenomenon**. The structure of the video signal is closely tied to the available technology for capturing, processing, transmitting and displaying images in the pre-digital era.

The video image is structured as a rectangular pattern of parallel lines, which are scanned continuously. The resulting signal represents the brightness at that point of the scan line. To record colour, three such scans are made simultaneously through appropriate filters recording three values in the colour model in use. The three values may be kept as separate component signals or combined as a composite signal.

The situation is further complicated by structuring the image as a frame made up of two **interlaced fields** of alternating sets of lines. The odd-numbered lines form the one field, and the even-numbered lines form the second. Each field is

---

[32] http://digitalpreservation.ssl.co.uk/restoration/

shown alternately when the image is presented. This has the effect of reducing the perceived flicker in the image as each field only requires half the bandwidth of the full frame and so twice as many fields as frames can be transmitted in a given time interval.

The quality of image that can be carried by such video signals has been closely linked to the **bandwidth** of the technology available to transmit and process them. This bandwidth has to be shared between the number of scan lines, the rate at which they are scanned and the range of chrominance (colour) and luminance measured at each sample. The historical development of broadcast television and professional and consumer video has seen a steady improvement in the quality of image that can be delivered to the viewer. However, as we see below, the quality of video images as delivered to the general viewer, whether by broadcast, video recording and even DVD, has always been quite heavily restricted by the available technology.

The number of scan lines and the mechanisms for recording colour within the restricted bandwidth available has led to a **variety of standards** used for analogue colour television including NTSC, PAL and SECAM.

Recommendation 601 (or *CCIR 601)* or ITU-R BT.601[33] [34], to give it its official name, introduced in 1980 by the International Telecommunication Union (ITU) (formerly known as CCIR (Comité Consultatif International pour la Radio)), forms the **basis for all modern digital TV** standards. It proposed an approach that allowed digital systems to break free of the constraints imposed by the legacy systems for analogue colour TV (such as NTSC, PAL and SECAM) and allowed manufacturers to produce digital TV equipment that would work on either 525-line or 625-line signals.

ITU-R BT.601 describes methods for sampling both luminance - the brightness in the video image - and chrominance - the colour information - and a colour **encoding** system in which the data for both are expressed. It also includes data for the horizontal and vertical sync and for the blanking intervals. It describes both interlaced and non-interlaced (often called **progressive**) formats.

There are many and detailed **differences** between the real-time world of **video signals**, even digitised, and the simple model of **digital image** handling. An extensive and sometimes amusing account from the point of view of a programmer of video software is given in C Pirazzi, *The Lurker's Guide to Video*[35]

Digitising a video signal involves sampling the signal at an appropriate rate and bit-depth; however, this will not produce the simple raster image described in our model of digital images. Unlike a film camera or a digital camera that 'grabs' the image at an instant in time, the video signal describes a potentially moving

---

33 International Telecommunication Union *BT.601-6 Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios*, ITU 2007

34 http://www.ebu.ch/fr/technical/trev/trev_304-rec601_wood.pdf

35 http://lurkertech.com/lg/

subject over a period of 1/25th of a second (PAL, SECAM) or 1/30th of a second (NTSC). Furthermore, the digitised video signal will preserve the field and frame structure.

A digitised pair of fields can be combined as a raster image to record a frame. However this will result in adjacent scan lines recorded 1/50th of a second apart (1/60th sec for NTSC) resulting in discontinuities in the recorded image. Recording the origin of the digital image as a pair of fields allows us the potential to play back the image as a pair of fields. If the image were displayed on a CRT monitor, the resulting experience would be authentic at the field level. However as CRT displays are replaced by digital monitors which display the frame all at once rather than by a raster scan, it will be increasingly difficult to recreate the original experience of viewing the video. By preserving as metadata that the image was originally interlaced the opportunity to exploit some future technology that can deliver an authentic experience at the field level can be maintained.

Interlaced	boolean	

We can also note that film has also been used for recording video by filming the video played back on a monitor. This would be recorded as part of the Provenance metadata associated with the digital object. By preserving this information the opportunity to use further processing to recreate a performance closer to that from the original video can be maintained.

4.5.4 Digital video

Increasingly archives are presented with Submission Information Packages (SIPs) in digital format. For the reasons discussed above (Chapter 1) these are likely not to be encoded in suitable preservation formats. Two principal issues arise:

Early digital video was recorded on tapes in **media-specific formats** such as DigiBeta. These cannot be held as normal files in a mass storage system and are not considered viable preservation formats. For preservation these need to be re-encoded in a suitable file format. This generally needs to be done as a matter of urgency as the equipment to play them is becoming obsolescent and, as the PrestoSpace investigations showed, the tapes themselves have very variable life expectancy even when kept in favourable environments.

Professional video formats and those used for digital cinema post-production preserve **frame integrity**. In order to support the editing function, each frame is encoded as a separate entity and any compression used is lossless. This means the encoding conforms most closely to our Significant Properties *sequence of images* model. Such a format may well be a suitable preservation format. Unfortunately video produced by consumer and semi-professional equipment will be encoded with **lossy compression** and hence needs re-encoding for preservation.

4.5.5 DVD

DVDs are not simple digital video objects. They have a structure that can accommodate, as well as the video, up to 8 tracks of digital audio, each with up to 8 channels and multiple subtitle tracks. They can also support various features for interaction including:

- instant search to title, chapter, music track, and timecode;
- instant rewind and fast forward;
- on screen menus;
- the ability to select different camera angles or viewpoints during playback;
- automatic branching of video for multiple story lines.

They can also include multilingual metadata such title name, album name, and creator details.

However not all discs support all features and there is considerable variation in the interpretation of the various standards applied to DVDs. The video is encoded in the lossy MPEG-2 encoding. They also incorporate digital rights management features which may interfere with the archiving mission.

Preservation of DVDs and related media-specific formats is considered outside the scope of this study.

4.6 The impact of encodings

4.6.1 Codecs

Moving image sequences have to be encoded as a bit-stream for transmission or storage by digital systems. The structure of this bit-stream is determined by the *codec*, the method used for encoding and subsequent decoding of the bit-stream for display.

We should note that a large proportion of the core, defining, technical characteristics of moving image sequences - their Significant Properties - are actually captured in the digitisation process or by the formats used for born-digital moving images. Although the encodings used may not be suitable for preservation purposes, the values of these Significant Properties have already been established as a result of the encoding used.

4.6.2 The tyranny of data rates

Recording digital images as a series of discrete frames with a *lossless* encoding can require large amounts of storage and the data rates required to display them can easily exceed what is available in a given environment. Professional video and broadcast organisations use the Serial Digital Interface (SDI), to provide sustained data rates in excess of 1 Gbit/s needed for digital cinema applications and HDTV. Uncompressed Standard Definition television requires a more modest 270 Mbit/s.

Commonly available 'Gigabit' Ethernet (1000base-T) can support sustained throughput up to 600 Mbit/s but achieving this is highly dependent on the traffic and can easily be much less. Hi-Speed USB (USB2.0) supports 480 Mbit/s. A DVD player can deliver 11 Mbit/s. ADSL promises 8 Mbit/s (downstream, ie to the viewer) but typically delivers in the order of 1 Mbit/s in the domestic situation. Digital terrestrial television typically uses 6 Mbits/s.

Clearly, general access to digital moving images cannot make direct use of the professional formats. Indeed most of us rarely experience video in this form. For any practical public access alternative formats requiring significantly lower data rates must be provided and some form of compression is needed.

4.6.3 Lossless and lossy compression

Lossless compression can reduce the data rate needed but not sufficiently in most cases. More significant reductions in data rate involve *lossy* compression techniques. These discard information while trying to minimise the visual impact of the loss. Being lossy, the compressed bit-stream cannot be used to regenerate the original bit-stream. However all digital video delivered to the public involves one or other lossy compression encoding technique.

Lossy encoding involves a reduction in the fidelity of the recorded image and hence to the authenticity of the performance. Hence, although it is not a property of the original source material, it is at the heart of the digital object and the nature of the encoding must be regarded as a Significant Property.

How to characterise the encoding?

- Encodings making use of lossless compression can simply be noted as such. The reduction in size of the bit-stream, the compression ratio, can be noted however, as the fidelity is not impaired, it is not a Significant Property.
- Encodings employing lossy compression result in a reduction of the size of the bit-stream at the cost of impairment in fidelity. The compression ratio can be easily measured:

  bit stream size (uncompressed) / bit stream size (compressed)

  The compression ration can provide a measure of the degree of impairment of fidelity, however it does not tell the full story. Different encodings produce different impairments in fidelity for the same reduction in size. There are standardised techniques[36] for the subjective assessment of the degree of impairment but these are not susceptible to automation. Nonetheless we can regard the combination of the compression ratio and the codec used as a significant factor having an influence on the fidelity of the performance.

The following Significant Properties in the rendering category relate to the encoding:

---

[36] International Telecommunication Union *BT.500-11 Methodology for the subjective assessment of the quality of television pictures*, ITU 2002

Lossless	boolean	
Compression ratio (if lossy)	ratio	dimensionless
Codec (if lossy)	name	enumerated set

Note that data rate or bit rate is often quoted as a measure of compression. However our view is that this is properly seen as a specification for the playback or transmission channel in order to enable it to play the decoded bit stream at the correct frame rate. The data rate needed can be calculated as:

data rate = bit stream size / duration

The compression ratio can be seen then to be the same as ratio of the data rates of the uncompressed and compressed bit streams:

compression ratio = data rate (uncompressed) / data rate (compressed)

Further investigation by subjective assessments may establish parameters associated with the encoding that could be candidate Significant Properties.

4.7 Significance

The InSPect project has proposed the use of a set of performance indicators identify the relative significance of the properties from a digital preservation point of view:

Numeric Value	Summary	Description
10	Essential and unchanged	Removal or damage to the property is likely to result in the inability to use or reproduce the message.
07-09	Essential. Some variation allowed.	The property should be maintained to recreate the Message. However, the value assigned to the property may be changed to some degree, intentionally or unintentionally without significant effect on the recreation of the message.
04-06	Beneficial	The property is used in the Message and may be maintained. However, other properties exist that perform the same or similar purpose.
01-03	Minor contribution	The removal or damage to the property results in minor loss and does not contribute to the significant loss to the Message.
0	Not Applicable	The property is unimportant for the reproduction of the message, does not contribute to the semantic understanding, or use of the Message.

In attempting to apply these to digital moving images we make the following observations:

The Significant Properties in the **Content** category:

- Different channels may be of different intellectual value and of different interest to different audiences. Their significance will vary accordingly.
- If the work is truncated it will not be possible to reproduce the message in its entirety. In this respect the *length* is highly significant. However there are other approaches to ensuring the integrity of the record such as the use of checksums.

The Significant Properties in the **Context** category all enrich the understanding of the message but are unlikely to be critical to it.

Most of the Significant Properties in the **Rendering** category are captured, in the digitisation process or by the formats used for born-digital moving images. Assuming that the encoded bit-stream can be decoded and played successfully the Message will be conveyed to some degree. In some sense, and ignoring factors of the ambiance of the Performance, the process of encoding the moving image sequence has determined the Performance. Different encodings result in changes to the Rendering properties, most of which can take a range of values. This in turn will affect the fidelity of the performance, in extreme cases rendering it unintelligible.

Methods for subjective assessment of the fidelity of a performance of moving images are described in the International Telecommunication Union's *BT.500-11 Methodology for the subjective assessment of the quality of television pictures*. For instance using the DSIS (Double Stimulus Impairment Scale) method, the subject is presented with an unimpaired reference video, then with the same video impaired. They are asked to rate the degree of impairment of the second video on a scale from "impairments are imperceptible" to "impairments are very annoying".

The techniques of BT.500 come from the television world and may need adaptation to be applicable to the much wider range of viewing environments used for digital moving images in general. An investigation into these and related approaches for assessing the fidelity of a performance could provide a means of assigning the relative significance of properties for moving images.

5 The Archival Information Package (AIP) for Moving Images

To be supplied

6 Content & Intent in Moving Images

6.1 Creation, Use and Significant Properties

A comprehensive approach to digital preservation requires a wider range of considerations to be taken into account than have been the concern of this particular study. We have already pointed out that we have deliberately focused on certain parts of the life cycle of digital moving images and within this our main interest has been in archiving requirements. Considerations relating either to the original creation of the objects or their future use have largely remained in the background. One particular reason for this is that moving images are more likely to be edited and otherwise modified than some of the other forms of digital object and, because of the need for compression in distribution, there is a sharper distinction than with other digital objects between the archiving or preservation format of moving images and their use formats. We have therefore focused on the preservation of core technical features. However, there are considerations relating especially to the creation of moving images that are of concern for a discussion of Significant Properties.

We saw that earlier work on Significant Properties, and especially the CAMiLEON project, extended the concept to include stylistic and aesthetic features and that these were seen to be tied to the creation and use of the objects concerned.

Our discussions with film archivists, in particular, has suggested that this is an issue of considerable importance in the preservation of moving images, where the intentions of the original creators, particularly in relation to the visual effects created are a matter of great concern. Often these issues arise in the context of film restoration, where there is an attempt to get back to a situation prior to damage being inflicted and where original intention is clearly an important consideration that can provide guidance to the people undertaking the restoration. More generally, however, there is clearly a relationship between content and the intent of the creator of moving images. It is worthwhile putting this in a broader context.

6.2 The Implications of the InSPECT Performance Model

One of the important aspects of the Performance model, adopted by the InSPECT project, is that it accepts that change is inevitable - that the rendering of digital objects will give rise to different ‘performances’, especially as the object is migrated from one format to another.

Traditionally, the concept of ‘performance’ has been fairly closely associated with the notion of ‘interpretation’, particularly in relation to the performance of plays or music. It is interesting that, over a long period, music has been preserved in music scores and plays as a text and that it has been accepted that the performance of a play or a piece of music is open to a considerable degree of interpretation. No-body expects a play or music to be performed in exactly the same way as it was in a previous ‘production’. On the contrary, exploring different interpretations is considered to be the normal course of events and part of a living cultural tradition. Some art forms go even further. With music, at least, it may be said that a musician is attempting to rediscover the original intention of the composer, but some poets take a completely different view, arguing that once a poem is in print, the intentions of the writer are irrelevant. The poem stands as a work in its own right.

There is, then, in artistic work, a spectrum of opinion, or rather 'attitude', on the importance of the intentions of the original creator and it is worth asking why, in the area of film, intentions are regarded as being so significant. In the context of preservation and archiving, of course, most of the material is not of an artistic nature, but even with documentaries, news coverage and other film and television output, the intent of the creator, with respect to the style of presentation, is often regarded as a significant part of the message. This is reinforced by the instincts of archivists and the implicit models obtaining in the preservation of documents or objects in collections, where authenticity of the original is a major consideration and where the circumstances of the object's creation, including the known or inferred intentions of the creators, are given a lot of weight. We ought, however, to be aware that other models have been employed for centuries in parallel areas and at least ask, as we have done, why the creator's intentions are so significant in the case of moving images.

One of the advantages of posing this question is that it enables us to have some perspective on the degree of significance that can be attributed to factors relating to intention. We take it as read that a significant proportion of moving image archivists and of the user community regard the original intentions of the moving image creators as important in determining some of their Significant Properties. That is to say, that we have to take these intentions into account, when migrating digital moving images from one preservation format to another. The argument is that, if we do not do this, when the new version is rendered, it will be perceived to be different in some essential respects.

The fact is, however, that the degree of interpretation of moving image objects is considerably more restricted than is possible with a musical score or the script of a play - unless re-editing is undertaken, at which stage we are talking about a different object. Once the core technical properties have been captured, most of the intentional elements are already fixed, since most of the characteristics that make a difference to the look and feel of different moving image presentations are reducible to different technical characteristics. A 1920's film clearly looks and feels different from a film from the 2000's. But, if we leave aside the question of what a 1920's director might have done with modern technology and assume that the way that a 1920's film looked was how the director intended it to look, most, if not all, of the difference in the look and feel is down to the technical characteristics – image rate, colour gamut etc. If we define the technical Significant Properties correctly and identify a format that carries them, we capture most of the intentional elements.

In Chapter 4 we aired issues about re-creating the characteristics of contemporary display devices in order to re-create an authentic experience. The date of creation and information as to provenance are likely to provide enough information to establish the technology available at the time. If there were unusual circumstance these could be recorded as additional metadata.

There is still room for debate about whether or not particular digital formats are good enough, but our interest is in defining a set of Significant Properties and the considerations just set out suggest that the scope for extension of our core set is rather limited.

Nonetheless, there will be contextual considerations that are not captured automatically. The ambiance or environment in which a film or video was intended to be seen may have been a significant part of the creator’s intentions and this would include such considerations as whether a film was intended to be seen through a special pair of glasses or other devices. These external elements cannot be captured automatically. They can only be expressed through appropriate metadata.

Furthermore, as we have said before, there is some merit, if the available resources allow, in including all the Significant Properties in the metadata schema to be used in the preservation package.

Our conclusions, then, are that much of the ‘significant’ intent of creators of moving images can be captured by our standard set of Significant Properties and the appropriate formats. The Significant Properties and other elements that cannot be captured in this way have to be expressed in a metadata schema.

7 Conclusions and Recommendations

This chapter needs revision to reflect the current version of the text

7.1 Conclusions

It is worth repeating that the study set out with relatively modest ambitions. Its aim was to build on a substantial body of earlier work (notably that brought together in the 2006 Digital Moving Images and Sound Archiving Study) and to contribute further to the development of practical procedures for professionals engaged in digital preservation. A particular objective was to develop a schema for characterising the Significant Properties of digital moving images on the basis of the model developed by the InSPECT project.

However, although these ambitions have been modest in one sense, we believe that they are nonetheless important in contributing to a wider understanding of what is required with the preservation of digital moving images. A common theme in this area is the need to facilitate the effective participation of non-archival organisations in the preservation process. We believe that in our study we have been able to clarify a number of issues in terms that contribute to a common and more widespread understanding of processes and procedures and that in particular we have developed a vocabulary and framework that can be shared by the range of stakeholders.

There is now a considerable amount of moving image material in digital form – digitised film and video in a modern digital format and born-digital cinema and video. But, equipment and computer programs change over time and lose their ability to read older files and display the information. Obsolescence is thus a major problem and digitisation as such does not solve the problems of preservation.

The core issue of the study then is defining the characteristics and features of the original sequences of moving images that have to be preserved, as they are migrated through preservation formats over time. This is what is meant by defining the 'Significant Properties' of digital moving images.

Significant Properties can be 'captured at different levels. In a sense, they are captured in the original data object as a set of bit sequences. For them to be reproduced, however, the data have to be expressed in an accessible format. Choosing the right format is therefore a major consideration in preserving Significant Properties. At another level, there is then the capturing of Significant Properties in metadata structures. These not only allow for an explicit recording of the core technical Significant Properties. They also permit flexible extensions, to cover, for example, contextual information that may be of importance for specific objects. There are, of course, costs associated with the recording of metadata that have to be taken into account.

It is important that making the process of defining and preserving Significant Properties operational should be achieved by building on the professional practice of digital archives, in particular, within the framework of the Open Archival Information System (OAIS) Reference Model. A core concept within this model is that of the Archival Information Package (AIP), which defines the preservation information package that is the focus of archiving strategy.

In effect, we believe that our exercise has been one of making a contribution to the definition of a standard AIP for moving images.

As far as the management of a digital archive is concerned, this leaves complementary work to be undertaken in relation both to the Submission Information Package (SIP) – how objects for preservation are delivered to the archive by the producer – and the Dissemination Information Package (DIP) – the package that allows users access to the digital moving images held by the archive. In the first case we are often talking about how moving image sequences are delivered to the archive or intermediary for digitisation and, in the second, we are referring to the various forms in which the archive distributes material that it holds, taking into account current delivery channels.

The contribution that is being made to defining a standard AIP for moving images deliberately makes use of the framework provided by the InSPECT project. InSPECT's approach to the definition of the Significant Properties is being offered as a standard approach, especially for the UK Higher Education sector, for a number of different types of digital object. It does, however, require that a further set of conventions are observed in the formulation of the Significant Properties of moving images.

It is important at this point to say explicitly which types of moving image have been addressed. Our focus has been on the main types of digital moving image (held in file format) that of interest to digital archives. This includes digitised film, digital cinema and broadcast and other output held in digital video. Unfortunately, we were not able to give much consideration to medical and scientific imaging and have not commented on it and our consideration of digital video has been restricted to device-independent video and has excluded video in physical formats intended for consumer equipment, e.g., DVDs. These exclusions indicate areas for further investigation.

Bringing these various factors together has led us to express the core Significant Properties of moving images as follows:

Content	Number of image and audio streams; length;
Context	Title; creator; date of creation; provenance;
Rendering	Frame width and height; bit-depth; gamut; pixel aspect ratio; frame rate; compression ratio; codec
Structure	Relationship between audio and image streams, relationship between metadata and bit-streams; interlace;

Significant properties defined in this way then become the criteria for determining the preservation format to be used by archives.

We found it difficult to assign the measurements of significance suggested by the InSPECT project were useful in analysing the properties with respect to preservation of the AIP and also for discussing appropriate formats for access to the digital moving images, candidate DIPs.

In the course of examining Significant Properties, a major conclusion of the project was that a large proportion of the core, defining, technical characteristics of moving image sequences are actually captured in the digitisation process or by the formats used for born-digital moving images. This implies that ensuring that the Significant Properties of moving images are reproduced is, to an important extent, a matter of using the right format.

The main candidate formats have already been determined by the Library of Congress. They are as follows:

To be supplied

In developing an appropriate metadata schema, for the explicit recording of Significant Properties, we have supported the developing use of METS and suggested how METS might be further developed to accommodate the requirements of digital archiving, as we have outlined them. We believe that a flexible regime, such as that provided by METS, is necessary as a way of taking account of the broader type of Significant Property relating to stylistic and aesthetic features and especially those that arise from the intent of the creator of the moving image sequence.

7.2 Recommendations

Defining the Significant Properties of moving images in this study has placed the process within the general context of the management of digital archives. This has revealed the need for parallel developments in relation to the specification of the Submission Information Package (SIP) and the Dissemination Information Package (DIP) that determine respectively how digital objects are submitted for acceptance by the archive (perhaps initially for digitisation) and how they are made accessible for users. For effective digital management further progress is needed in relation to these matters.

Other recommendations that can be made that relate more specifically to the definition of Significant Properties:

- The use of the InSPECT data model would benefit from a joint effort to align the descriptions of Significant Properties used by the different Studies in applying the model;
- The development of interactive elements in broadcast media and elsewhere begin to bring in a degree of complexity that we believe can best be addressed by looking at interactives in general, in a similar exercise to that undertaken in this study;
- The issue of restoration to remove defects in the original recording technology or to repair damage needs further investigation to enable archives to develop an informed policy on the restoration techniques that become possible.
- Medical and scientific imaging raises a further set of issues, some of which require resolution of the approach to be taken with regard to interactives. They are, however, of particular importance in the higher and further education sector and need to be addressed in a further study.

Our study has made suggestions in terms of developing a coherent metadata structure for the flexible recording of Significant Properties and other key preservation information. We believe that the full development of the metadata schema will require resolution of issues that are beyond the scope of this study, with provision for all the elements that are required, for instance, to provide access (such as intellectual property rights information). Again the recommendation is that a general solution be found, incorporating the elements we have highlighted in a framework that covers all aspects of digital management.

It is important that developments in digital preservation in the UK relate effectively to those taking place elsewhere and not only in the English-speaking world. With European Union policies relating to Digital Libraries and the development of a common European Research Area in digital preservation beginning to show concrete results, there is a need to ensure that UK digital preservation continues to relate to developing European and global practice and takes advantage of the economies of scale available in such interaction. It will be particularly important to take up the results of the CASPAR and PLANETS projects, not least through the agency of the UK institutions that are active in them and we recommend that a watching brief be kept on them.

8 References

8.1 Citations

S Anderson et al, *Digital Moving Images and Sound Archiving Study*, AHDS 2006

M. Coyne et al, *The Significant Properties of Vector Images*, JISC 2007.

Dublin Core: ISO 15836 or ANSI/NISO Z39.85 –2007 : http://www.niso.org/standards/standard_detail.cfm?std_id=725

Margaret Hedstrom, Christopher A. Lee, *Significant Properties of Digital Objects: Definitions, Applications, Implications*, 2002.

G. Knight et al, *Framework for the Definition of Significant Properties*, AHDS 2008.

G. Knight, *Video Preservation Handbook*, AHDS 2005

G. Knight, *Audio Preservation Handbook*, AHDS 2005

International Telecommunication Union *BT.500-11 Methodology for the subjective assessment of the quality of television pictures*, ITU 2002

International Telecommunication Union BT.601-6 *Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios*, ITU 2007

METS: *The METS Reference Manual, Metadata Encoding and Transmission Standard: Primer and Reference Manual* http://www.loc.gov/standards/mets/METS%20Documentation%20final%20070930%20msw.pdf

Thibodeau, K. 2002. '*Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years'* in proceedings of The State of Digital Preservation: An International Perspective. Conference Proceedings. Washington. 2002.

OAIS: *ISO 14721:2003 Space data and information transfer systems – Open Archival Information System – Reference Model* http://public.ccsds.org/publications/archive/650x0b1.pdf

C. Pirazzi, *The Lurker's Guide to Video* http://lurkertech.com/lg/

PREMIS: http://www.oclc.org/research/projects/pmwg/premis-dd.pdf

The National Archives (TNA) and the UK Data Archive (UKDA) '*Assessment of UKDA and TNA Compliance with OAIS and METS Standards'* 2004.

Andrew Wilson, *Significant Properties Report*. 2007. Library of Congress, Sustainability of Digital Formats: Planning for Library of Congress Collections http://www.digitalpreservation.gov/formats/content/video%20preferences.shtml#prof

8.2 Projects

Cedars
http://www.jisc.ac.uk/whatwedo/programmes/programme%20rep%20pres/inspect.aspx

The Cedars Project Report (April 1998–March 2001), June 2001, p 14
http://www.leeds.ac.uk/cedars/admin/CedarsProjectReportToMar01.pdf

CAMiLEON http://www.si.umich.edu/CAMILEON/

InSPECT http://www.significantproperties.org.uk/

CASPAR: http://www.casparpreserves.eu/publications/deliverables

PLANETS: http://www.planets-project.eu/

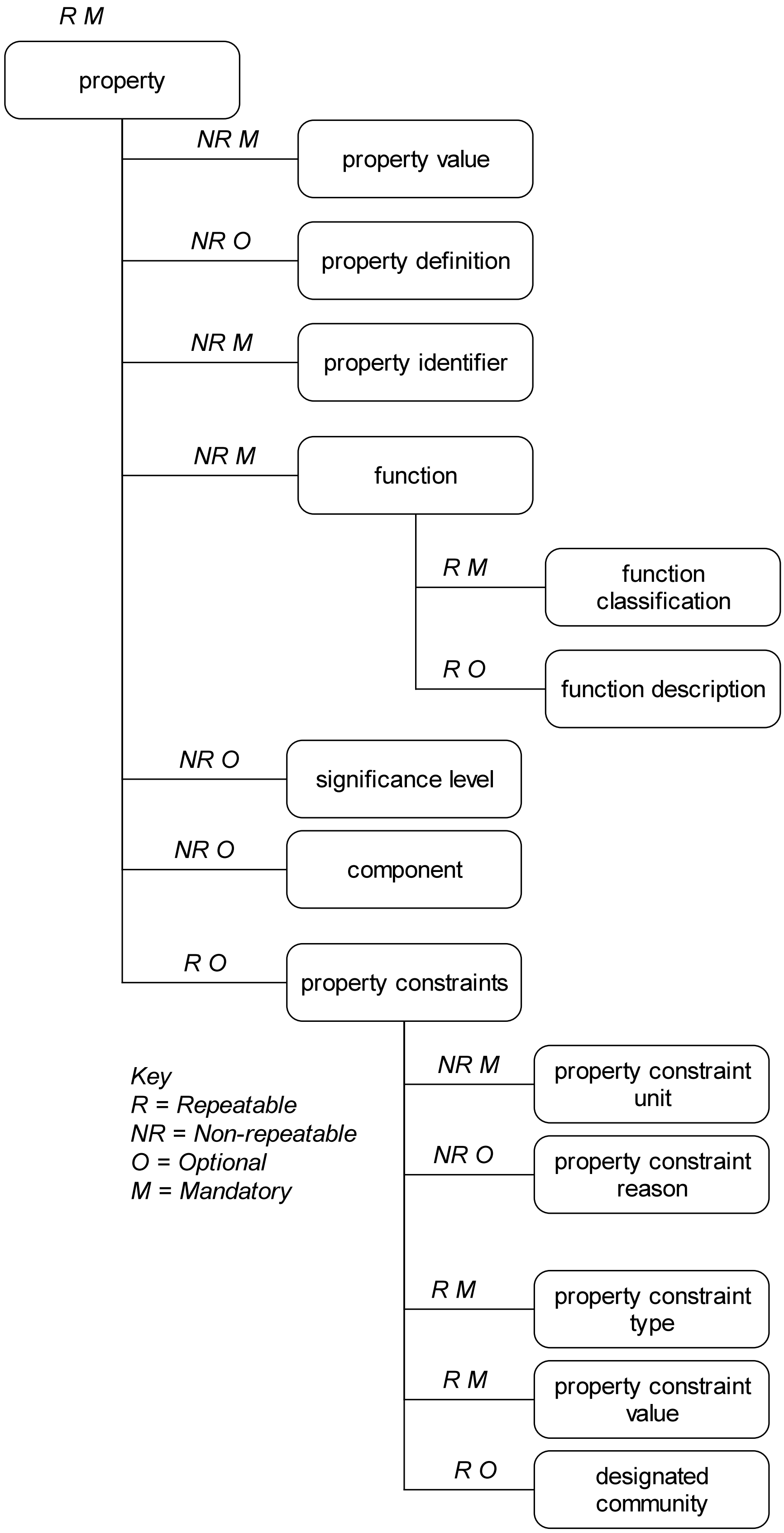
PrestoSpace: http://prestospace.org/, http://digitalpreservation.ssl.co.uk

8.3 Resources

The **PRONOM registry** is accessed through TNA's web site:
http://www.nationalarchives.gov.uk/aboutapps/PRONOM/tools.htm

JHOVE: http://hul.harvard.edu/jhove/index.html

Appendix A: InSPECT Data Model



The containers, sub-containers and elements illustrated above perform different functions necessary to identify, describe and measure the significant properties of a Record. The following definitions are used:

- *propertyValue:* The title of the property that indicates its purpose. The title should be unique to the Record or Component to which it is applied to avoid confusion and, if possible should remain consistent across similar file types.
- *propertyDefinition:* A formal statement that describes the purpose of the property. The definition provides a human-readable description of the property. It should be stored by an appropriate service provider, such as The National Archives' PRONOM and is not intended to be stored in the Record metadata itself
- *Property Identifier:* A machine-processable identifier to categorise each property.
- *function*
    - *functionClassification:* A controlled vocabulary that indicates the high-level function that the property performs in the Record.
    - *functionDescription:* A free text description of the function that the property performs
- *significanceLevel:* An assessment of the significance of the property to the recreation of the Record
- *component:* The sub-component of the record with which the property is associated
- *propertyConstraints*
    - *significantPropertyConstraintType:* An indicator of the type of constraint placed on the value of the property.
    - *significantpropertiesConstraintUnit:* The unit in which the value is measured
    - *significantPropertyConstraintValue:* The measured value of the constraint
    - *significantPropertyConstraintReason:* A free text box that may be used to explain the rationale for the constraint
    - *designatedCommunity:* Multiple value constraints may be applied to different user groups in the Designated Community. The designatedCommunity value allows the archive to declare the acceptable property constraints that may be tailored for different user types. Possible examples of two designated communities are 'archive' for institutions performing preservation and 'dissemination' for academics and other users.

Source : Gareth Knight, Framework for the Definition of Significant Properties, AHDS 2008.

Appendix B: Table of Significant Properties

property value	component	property definition	function classification	function description	significance level	constraint type	constraint reason	constraint unit	datatype
imageStreams	video	The number of distinct video channels in the digital object	content / rendering					Channels	Positive integer
audioStreams	audio	The number of distinct audio channels in the digital object	content / rendering					Channels	Positive integer
length	video	The number of frames in the recording	content					Frames	Positive integer
width	frame	Width of the image frame	rendering	Contributes to spatial resolution				Pixels	Positive integer
height	frame	Height of the image frame	rendering	Contributes to spatial resolution				Pixels	Positive integer
bitDepth	pixel	The amount of data recorded for each pixel.	rendering	Defines the colour resolution				Bits	Positive integer
colourModel	pixel	The colour model used when recording the pixel value	rendering	Contributes to the specification of the Gamut				Enumerated list	
colourSpace	pixel	The colour space used when recording the pixel value	rendering	Contributes to the specification of the Gamut				Enumerated list	
pixelAspectRatio	pixel	Aspect ratio of the pixels making up the image	rendering	(Together with *width* and *height)* determines the aspect ration of the frame				Ratio	Ratio
frameRate	video	The number of frames to be displayed in a given time interval	rendering	Determines the rate at which the frames are displayed				Frames/second	
lossless	video	Whether the encoding is lossless	rendering	Indicates the potential for loss of fidelity through encoding				Enumerated list	Name

property value	component	property definition	function classification	function description	significance level	constraint type	constraint reason	constraint unit	datatype
compressionRatio	video	Measures the degree of compression employed by the encoding	rendering	If a lossy encoding is used this contributes to a measure of impairment of fidelity				Ratio	Ratio
codec	video	Identifies the codec used for encoding	rendering	If a lossy encoding is used this contributes to a measure of impairment of fidelity				Enumerated list	Name
interlace	frame	Whether the source was interlaced video	structure / context					Boolean	Boolean
metadata		The existence of metadata associated with the digital object.	context			equality	Indicates the number of metadata elements in the Record	Elements	

Table B1: Significant properties of digital video (excluding properties intrinsic to audio aspects)