The Arts and Humanities Data Service (AHDS) is a national service funded by the Joint Information Systems Committee (JISC) of the Higher and Further Education Funding Councils and the Arts and Humanities Research Council to collect, preserve and promote the electronic resources which result from research and teaching in the arts and humanities.

Digital Images Archiving Study

Authors:	Sheila Anderson Mike Pringle Mick Eadie Tony Austin Andrew Wilson Malcolm Polfreman
Date:	March 2006
Version:	Final draft
Document Name:	Digital Images Archiving Study
Notes:	Limited circulation

Acknowledgements
Many thanks to those who contributed to this Archiving Study by participation in interviews and through responding to the survey promoted online through the AHDS Visual Arts.

Digital Image Archiving Study: Table of Contents

1. Summary and Recommendations

1.1 Background and Focus

A growth in digital imaging technologies has meant that a vast number of digital images are produced every year. The scholarly community has made particular use of this new content, and is now strongly reliant on access to digital image resources. Digitisation of collections has allowed a radical shift in the manner in which libraries and cultural organisations can deliver their collections, enabling them to provide round the clock online access to multiple users from anywhere in the world.

While many images are "born-digital", in particular vector images, vast quantities of digital images have been digitised. Over the last ten years, costly digitisation projects have been carried out in many HE/FE institutions. If this investment in digitisation is to pay off, then strategies for the long-term preservation and access of digital image files must be guaranteed. Digitisation projects need to consider the fact that digital obsolescence will have a negative impact on their digitised collections, in the same way that physical collections are subject to threats over time, and that the decisions they make at the point of digitisation or capture is likely to affect what they can do in the future.

Management and preservation requirements for digital materials are fundamentally different from analogue materials. Digital materials can be created using a wide range of technologies and formats, whether born digital or digital surrogates of existing analogue materials. They can be described and documented in a variety of ways – or not at all. They are subject to both physical deterioration and technical obsolescence. More than one copy can be easily and simply created. Access may be provided through more than one point, and may be distributed. All these factors will impinge upon the approach taken to their management and long-term preservation.

These differences present the curators of digital materials with some fundamental challenges. The way in which materials are created, particularly the technologies used, will determine how conducive to long-term preservation the materials are, and will present varied challenges to curators charged with the subsequent management and preservation of the materials. Curators will need adequate metadata about the resource if they are to successfully manage, preserve and make the materials accessible. Multiple copies may also imply multiple versions – the digital resource curator must somehow ensure the integrity and authenticity of the resource. They must be aware of changing technologies and fragility of media and take these into consideration from an early stage in the ingest process.

Access and preservation of digital content are of course closely linked. The conversion of binary data into meaningful information relies on a complex chain of hardware, software and formats, all of which are subject to on-going technological change. Consequently, providing long-term access to digital content inevitably involves the challenges of digital preservation. Content in a digital repository must be protected from the problems of data corruption and technological obsolescence, and the authenticity of the content must be ensured in some way.

Jones and Beagrie define digital preservation as: "…the series of managed activities necessary to ensure continued access to digital materials for as long as necessary.

Digital preservation….refers to all the actions required to maintain access to digital materials beyond the limits of media failure or technological change."[1]

All this suggests that digital curation and preservation requires a more pro-active approach beginning at an earlier stage in the material's lifecycle than would traditionally be the case with analogue materials. Within the digital preservation community, the concept of the life-cycle (or continuum as it is sometimes called) management of digital resources has emerged to describe and document the active management processes that need to take place, and the key decision making and intervention points along the continuum. The life-cycle concept has been incorporated into OAIS Reference Model, now adopted as an ISO standard for digital preservation. The OAIS model is proving a strong foundation for the development of digital archiving projects and services, and is increasingly being implemented by digital libraries and archives, including the AHDS.

However, despite these developments, the difficulty for those undertaking preservation or with responsibility for providing access in the long term to digital is the lack of practical advice, and of robust tools and mature techniques for digital preservation. A number of digital preservation strategies have been proposed, but there is no definitive approach to the problem of maintaining digital content across multiple generations of technology. Unfortunately, information on the likely costs, possible limitations and long-term sustainability of different strategies is far from complete – partly it must be said, for the very valid reason that no one has yet had the time to gain the experience needed to answer these questions.

The work undertaken for this report sets out to overcome some of these limitations, and to provide a firm foundation for the JISC in its future decision making. The report should also act as a valuable resource for those creating, managing, curating, providing access to and preserving digital images.

The definition of a digital image used in this study is as follows:

"Digital (still) images are non-moving representations of visual information"

That is, still images that convey their meaning in visual terms, e.g. photographs, posters, diagrams, drawings. The study considers both the familiar raster image and the perhaps less well known vector image. The former include the products of digital photography and scanning with file formats such as TIFF and JPEG. The latter is considered less when thinking of digital images, but a large volume of digital content is created including maps, drawings, and the almost ubiquitous PDF file.

Both can be said to be geometric or spatial but any similarities end there. Raster (or bitmapped) images are grid-based with information being held about each point or pixel within the grid, whereas vector images have information about any number X, Y, Z spatially defined coordinates and are made up of scalable objects—lines, curves, and shapes—defined in mathematical terms, often with typographic insertions.

The report focuses on a number of practical issues related to the preservation of images. First among these is the nature of digital images themselves – their content, format, size, metadata requirements, and potential preservation methods to ensure long term access and use. The report also addresses user community requirements,

[1] Jones, M and Beagrie, N, 'Preservation Management of Digital Materials: a Handbook', British Library, London, 2002

both for those creating digital images and for subsequent users of digital images. The final section endeavours to bring many of these issues together focusing upon a life-cycle model and some discussion of costs.

What is surprising is the lack of focused research dedicated to researching the preservation of digital images, and the lack of practical advice, and of robust tools and mature techniques for digital preservation, particularly for vector images. Research tends to be theoretically-based, and although testbed projects have begun to emerge, but overall research tends to be general in nature rather than say, investigating TIFF as an image preservation format. Work specific to digital image preservation is rare, although the programme to develop a decision support framework for the preservation of a range of content types currently underway at the Library of Congress is providing useful, focused research and practical guidance. **We recommend that JISC seriously consider funding work to complement and add to this programme in collaboration with the LoC, format registry work such as PRONOM and the Global Digital Format Registry.** Further work of this kind to identify the complexities, the risk factors and risk assessment processes, would be of enormous value to the preservation community, and would facilitate informed decision making at all stages of the digital life-cycle.

1.2 User Requirements

It is now widely understood that the development of digital systems, like any non-digital system, should be led by an understanding of the needs of those people who will use the system. This report gives great weight to what we have termed end-users, but it also includes within this remit those who are engaged with digitising or creating born digital images – the so-called start-users. The report argues that consideration must be given in the digital image preservation process to the users of those images, and must also engage with the creators of image collections wherever possible.

The research undertaken for this report highlights that user concerns are dominated by worries, some founded and some speculative, relating to, first and foremost, access and the metadata required to facilitate access, followed closely by the difficult issue of copyright/IPR, and the more transient nature of quality and changes in technology. We would therefore argue that it is not just the technology and systems that must be considered but also the connection between *technological facility* and *human factors*. Just as the different image formats develop, change or migrate into emerging formats, so a system must adapt and grow in response to issues relating to change of use, copyright, wider accessibility, and even metadata, within the communities that the system is designed to serve.

The concept of the lifecycle model developed for the report identifies that preservation starts with the data creation process, and must include regular monitoring and response to user needs and requirements if the right preservation decisions are to be made. For example, it is pointless for an image repository to migrate its content to one format, when the community doesn't use that format or have access to software that accepts that format. Similarly, an archive might concentrate on creating technical metadata that fits it preservation needs, but if that isn't accompanied by metadata that enables discovery and use, then it is questionable if the repository is fulfilling its remit.

Furthermore, preservation decisions must be based on the content and nature of the image and different preservation decisions and actions may be taken depending on these factors. For example, if the content of an image is unique (an original artwork for example, or a set of architectural drawings) then the approach to preservation may well be different to something that is more common, or has less value in terms of use (although assessing value is very difficult). Similarly, the nature of the image will affect preservation approaches and actions. For example, if precise colour depth and accuracy are not essential characteristics (as may be the case with many vector graphics) then the preservation approach is likely to reflect this. The concept of fitness for purpose is crucial to these ideas, and should be factored into the preservation life-cycle approach.

There seems to be a general assumption that we must preserve all digital content and every digital image that has been created. Indeed, that assumption was reflected in the ITT for this research. The concept of fitness for purpose brings that idea into question. What is fit for purpose now, may not remain so in the future. The need for different types of content may change over time for example, and if past experience is anything to go by standards of digitisation and image capture will continue to improve well beyond what is acceptable now. This may well render the nature of the object obsolete in terms of quality and/or functionality even though it can still be rendered and used in the technological sense.

Selection of images for preservation, and their subsequent retention require more attention than seems to have been given to them thus far. We would recommend that JISC investigate issues of selection and retention with the idea of fitness of purpose in mind. This should include assessing where the costs of digitisation of images lie. It may be that the most costly part is the creation of metadata rather than the capture of the digital image. In that case, JISC might recommend or require properly managed preservation of the metadata and surrounding information, alongside a programme to re-digitise the image content at such time as the nature of the object no longer meets the needs of the user communities.

However, if the cost of digitising the object is high, and the image content is born digital, or digitised from a fragile or very rare and hard to access analogue, then a different approach is necessary, where the image is created to the highest possible standards, and properly preserved into the future, alongside the metadata. This report recommends that JISC moves away from a 'one-size-fits-all' approach to preservation and investigates a range of different approaches that reflect the requirements for the future community of users, as well as the more prosaic technological requirements.

To address these issues JISC is recommended is undertake further research into the selection and retention of digital images, the associated costs of re-digitising against the costs of preserving, and how it might implement an on-going assessment process of user needs and fitness for purpose of digital image content.

JISC should require that all digital image content which it funds should be created to the highest standards possible to ensure usability for as long as possible.

JISC may also wish to consider:

- **Capturing and analysing user and usage data to monitoring changes in practice that can feed back into the development of systems**
- **Requiring image collections which it funds to conform to agreed standards, and providing monitoring services to ensure compliance**

1.3 Properties and File Formats

Access and preservation of digital content are of course closely linked. The conversion of binary data into meaningful information relies on a complex chain of hardware, software and formats, all of which are subject to on-going technological change. Consequently, providing long-term access to digital content inevitably involves the challenges of digital preservation. Content in a digital repository must be protected from the problems of data corruption and technological obsolescence, and the authenticity of the content must be ensured in some way.

However, as the chapter on user requirements demonstrates digital content must also be fit for purpose over time. This means that attention must be paid to establishing the significant properties and addressing the issue of file formats. To enable decision making about preservation this report has used the following three-tiered representation:

1. Preservation of the bit stream (basic sequences of binary digits) that ultimately represent the information stored in any digital resource
2. Preservation of the *information content* (images, sounds, moving images etc.) stored as bits and defined by a logical data model, embodied in a file or media format.
3. Preservation of the *experience* (speed, layout, display device, input device characteristics etc.) of interacting with the information content

All three still present challenges for digital preservation.

File (or data) formats define the rules used by application software to convert bits (the fundamental unit of digital data) into meaningful information that can be viewed and manipulated by a user. Most application software developers produce file format documentation for the formats they design and develop. Not all of them make this documentation available and even if they do, it is not always accurate (see Lawrence *et al.*, 2000, pp. 13-15 for examples of attempts to retrieve the Lotus 1-2-3 and TIFF file formats from their developers.).

Based on the availability and stability of the format specification, file formats can be classified as proprietary, open or standard formats. ***Proprietary file formats*** are not public and are developed and maintained by software producers. Larger software producers may sometimes publish their format specifications (PAS – Publicly Available Specification) or several firms may join together in a consortium to define interface standards so that they can develop mutually compatible products. These are called ***open or public file formats***. Some file formats are developed to become international standards (***standard file formats***) which are then public and fixed or stable until the next release of the standard. It is not unusual that software companies produce their own modified, proprietary, versions of standard file formats – these will be based on standards, but will have extensions that are proprietary and generally not public (e.g., Microsoft's version of XML). Many proprietary formats are, nevertheless, widely used and provide extensive compatibility with application software – these formats are often classified as *de facto* standards (cf. DLM 1997, pp. 50-52).

Successful and cheap long-term preservation of a digital file depends on the openness, level of standardisation and compatibility with other software products of the file format. Without a format specification the vital rendering tools that enable the use of digital files over longer time cannot be developed. Reverse engineering of software or the digital objects themselves can provide some answers, although legal constraints may well prevent this kind of action. Even where reverse engineering is possible, without any file format documentation, the process is likely to be too laborious and expensive (Leeds, 2003, p. 4).

The preservation risks associated with file formats are mostly related to loss of data and cost. Both migration and emulation — the two best digital preservation strategies currently in use — rely on file format specification being known and accurate. If it is not, then preservation actions risk introducing distortion, loss of quality or data, or not being able to render the file usable at all. The risk management of file formats for preservation has to account for all these considerations.

1.3.i Vector Images

This report identifies significant challenges for the preservation of vector images. The file formats and software applications used to create vector images are numerous, are frequently proprietary, and many generate binary files which can be problematic for longer term preservation. Fortunately many, but not all, allow for export as structured ASCII text which is generally accepted as being the most stable of formats.

However, an ASCII export is not enough in itself. Its meaning has to be openly documented in order that the data can be re-constructed for use in another software application. This can be expensive and time consuming as it usually involves a great deal of human effort to capture the correct information. It may also place a significant burden on subsequent users of the data as a relatively high level of expertise is likely to be required to manipulate and import the data. Moreover, the risk over the longer term of degradation and loss of functionality associated with this approach is unknown. We therefore **recommend that JISC undertakes further research and experimentation to test the suitability of this approach and to assess the risk of degradation and distortion that may occur during this process.**

Whist this would make significant strides towards ensuring the long term preservation of much vector data there is also a need to work alongside the recommended research to maintain information on the most suitable software to render these data in the future. Therefore t**his report also recommends the creation of a registry of common file formats which is updated regularly with comments on new and emerging formats and their suitability for use.** This goes beyond the usual format registry which advises on suitable preservation formats and provides technical details, to include information useful to users when making decisions about which formats and software to use. There is a requirement with many of these formats to capture additional information that may not necessarily appear in the structured metadata. For example, CAD files require fairly detailed information about the methodology and tools used to capture the image. If this additional information is not captured it can affect the long-term understandability of the data and thus degrade the value of the image. **JISC is also recommended to investigate the ISO StEP standard for the exchange of CAD files.**

Issues also exist with PDF files. Although Adobe is developing PDF+A as a standard to archive PDF files, it isn't yet certain how effective this will be. We therefore **recommend that JISC investigates the viability of the PDF+A standard and works with the vendor to ensure it meets preservation and re-use requirements.**

The issues with the preservation of vector image data identified throughout this report make it advisable that JISC regards further research and development in this area a priority for future action.

1.3.ii Raster Images

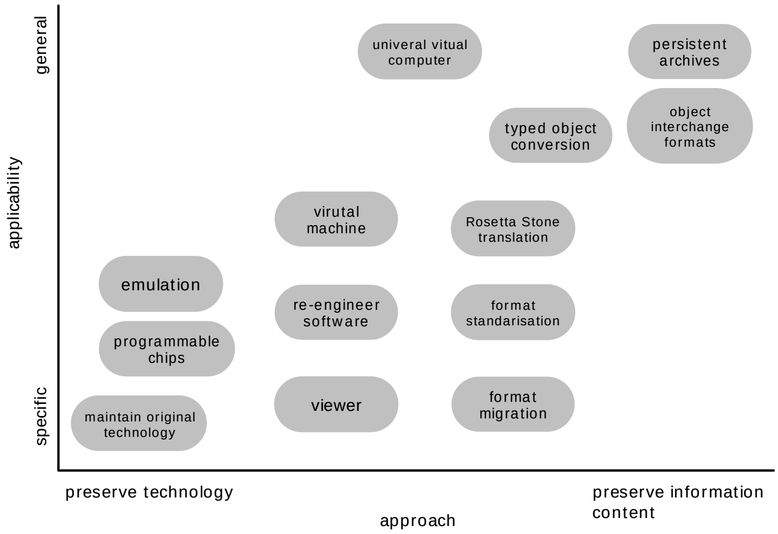
The situation with raster images is less complex. TIFF is widely regarded as a de-facto preservation standard and the emergence of JPEG2000 may well prove to be an acceptable standard for preservation. However, it is still worrying that little research has been undertaken to test the use of TIFF and this report would recommend that such research into both TIFF and JPEG2000 be commissioned. It is important in this context that this research involves the computer science community working alongside curatorial staff. It would also be appropriate to define a common set of significant properties to be adhered to by JISC projects creating digital raster images.

Specific recommendations include:
Colour issues need careful recording and working in RGB is recommended

There is a need to undertake further work to investigate the variability of TIFF and the consequences of not including additional functionality that is available in some revisions. Similarly PNG may offer potential but has little industry support and has the same problem of additional functionality. DNG has been developed by adobe to work with RAW files, primarily from digital cameras and may become a standard archive format. **We recommend that JISC undertake a tech watch / File format watch in this area, and that JISC engages with the industry in order to influence developments in this area, and to ensure representation of the JISC community**

1.4 Preservation Methods

Digital image data in whichever format remains useful only for as long as it can be correctly *rendered* (displayed, played-back, interacted with) into meaningful content such as text, images and video clips. The process of rendering is performed by a complex mix of hardware and software, which is subject to rapid obsolescence. As a rule of thumb, it is reasonable to predict that current hardware and software will be able to correctly render a file for around ten years after its creation. By the end of this period, repositories need to have adopted a more active preservation strategy than simply preserving the bit stream of the file if they are to maintain access to information content held in the file. Either old data must be altered to operate in a new technical environment (migration, format standardisation) or the new environment must be modified so that it can render the old data (emulation, virtual computers). Within these two broad approaches there are many different techniques:



Digital Preservation Strategies (based on Thibodeau, 2002)[2]

This report discusses the relative merits of technology preservation, migration, emulation, migration on request, and the universal virtual computer. Format migration and format standardisation appear to be the most common approaches for both vector and raster images, although significant problems arise when dealing with vector image files.

Emulation, migration on request, and UVC all pose significant problems and the injection of major programmes (and cash) to create the necessary migration and rendering tools, and decoders. At the present time it seems unlikely that these methods will be viable for repositories and unless and until the tools exist to undertake these processes they are likely to remain only in testbed environments rather than as working methods in actual repositories. The difficulty with emulation, migration on request and UVC approaches is the technical effort required to create the tasks to emulate or migrate on request. Whilst in theory all these approaches have merit, and look very attractive in some cases, the fundamental problem arises that unless there is a coordinated effort to create the tools (and as opensource so all can use) to archive the tools, and to make them widely and freely available through tools registries, it is difficult to see how they can be fully relied upon and implemented within working repositories.

We recommend a watching brief should be kept on developments in this area through the Digital Curation Centre and should suitable tools emerge then funding for developmental implementation in working repositories might be funded. It seems therefore, that for the time being, either format migration and format standardisation should be implemented. **JISC may also wish to investigate the feasibility of developing a consortium of commercial players, repository and archives technicians, and other interested parties to undertake a programme of developing these tools for the most used image formats.** However, experience with the KB suggests that this may well prove to be problematic and time-consuming, and should only be undertaken after careful consideration and consultation.

In terms of preservation practice, we would recommend that curators archive a copy of the original file in the event that emulation or other techniques become viable in

[2] 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.

the future. Retention of source bitstreams is a necessary component of any preservation approach to safeguard against migration errors and choices of preservation formats which might prove to be incorrect over time. It is also important to bear in mind that preservation formats are not necessarily distribution / delivery formats (although dissemination / delivery versions of digital images may be in the same file format as the preservation copies), and very careful consideration needs to be given to selecting the 'master copy' for preservation.

Some other points need to be made:

- There is no **single** best way to preserve any digital resource;
- Decisions about preservation approaches depend on resources available, current and future use of the resources, and the cultural/historical/social/legal significance of the resources;
- Decisions made about the recommended preservation formats for individual resources can change over time.

1.4.i Raster Images

At the present time the most reliable form of preservation for raster images is the format migration approach. This report recommends this as the default but archives and repositories are strongly recommended to also keep copies of images in their original format as create/deposited. This would ensure that should other preservation methods such as emulation or UVC became viable options in the future these can be adapted as necessary. The use of uncompressed TIFF version 6 is the best strategy at the current time, but a watching brief should be maintained on JPEG2000 as an emerging preservation format. Should as seems likely, industry support JPEC2000 format grow then this may well sit alongside TIFF, as the format of choice. A similar watching brief should be conducted for PNG and DNG.

1.4.ii Vector Images

The situation with regards to Vector files is much more complicated, primarily due to the large number of vector image software applications in use, and the fact that may of these use proprietary binary formats. These difficulties are compounded by the lack of non-proprietary open formats. Whilst support for the OGC data formats is growing, it is not yet widespread. JISC should work with the OGC community to encourage development in this area and to promote more widespread uptake and use across the industry.

Recommending a suitable preservation format is much more difficult for vector images and the approach will differ according to the purpose of the original data and the approaches listed in this report should be regarded as interim suggestions only and should not be relied upon for preservation of such files in excess of 10 years. The lowest common denominator approach would be to export the data as structured ASCII text with the creation of additional documentation that describes the meaning of the ASCII text structure and how to recompose it for use. There is an added complication in that this is not a particularly user friendly approach and does not encourage re-use of these materials. It is also a fairly labour intensive approach and hence adds to the costs. Unfortunately it is difficult at this stage to recommend any other approach and **JISC is strongly recommended to conduct more research into this area as a matter of priority**.

1.5 Metadata

This report discusses three types of metadata, all of which must be considered in the preservation of digital images. These are technical metadata, management (administrative) metadata, and discovery/use metadata. Technical metadata is necessary to describe the physical attributes of digital objects and is particularly important for preservation and rendering. Management metadata is essential to ensure authenticity, rights, ownership and provenance are properly addressed. Discovery and use metadata is essential to ensure future use of digital objects – being able to locate, access and use digital content in the long-term is arguably the raison d'être of preservation.

For raster images a range of metadata standards exist that might be used to capture and structure metadata, starting with basic Dublin Core for discovery and use, through more comprehensive standards such as VRA Core 3.0, PREMIS, and NISO Z39.87. Each has its strengths and weaknesses. For example VRA Core is very well suited for describing digital representations as it also provides a facility to describe the original work and any surrogate analogue such as a slide. However, it does not include a comprehensive set of technical metadata element and cannot therefore meet the requirements for preservation. Z39.87 provides a comprehensive metadata set. However, it may prove too comprehensive to enable wide take-up. The burden placed upon creators and curators of digital images to create metadata to this standard may be too high. Similarly, the costs involved may also prove too high.

This report therefore has taken a somewhat pragmatic approach and recommends a generic standard that maps across these more comprehensive standards. A convergence between the generic elements of PREMIS and the mostly mandatory file information section of 239.87 is therefore proposed. We also recommend the use of METS as a flexible wrapper within which the different element sets could be combined and preserved.

However, vector image formats present more of a problem. There are a number of emerging standards at different levels of development and acceptance for CAD, GIS, URML and SVG. **We therefore recommend that JISC should undertake a study to identify the common properties applicable to most (if not all) vector formats, and to work with the standards community to achieve this aim.**

Further recommendations in this area include the following:

A pragmatic study should be undertaken to identify common properties applicable to all or most vector formats.

Efforts should be made to create or modify metadata extraction tools (e.g. JHOVE, NLNZ Metadata Extractor Tool) so that they generate standardized and compatible metadata compliant with PREMIS and key format-specific schemas, such as Z39.87/MIX. Such tools should adopt elements from established namespaces and avoid the use of application-specific schemas.

Work should be performed to improve the tools used for metadata extraction to simplify the process of batch processing a large number of files in a collection and generating distinct XML records for each object. To maintain the validity of the original collection, XML records should be output to a directory structure that mirrors the original.

Further work should be undertaken to extend the range of common file formats recognized by metadata extraction tools.

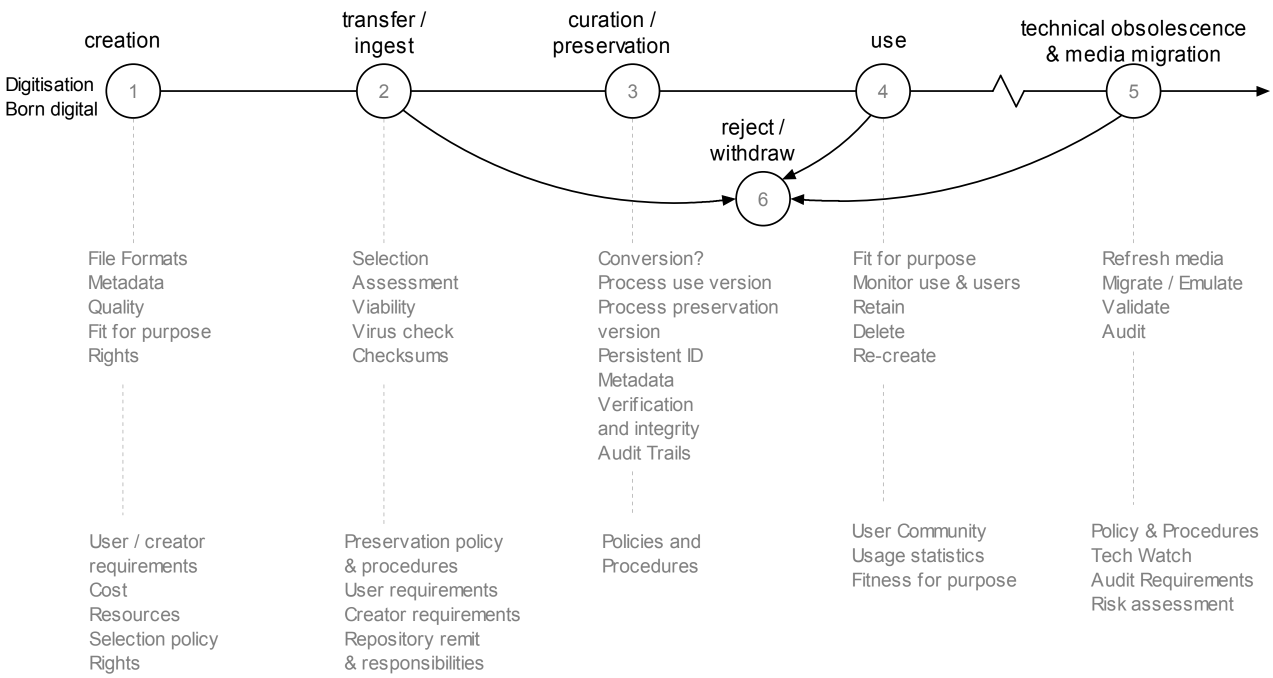
Work should be undertaken to assess the feasibility of automatic extraction of subject keywords using pattern recognition software.

Scenarios for integrating the manual and automatic production of metadata for digital still images and possible workflow issues should be investigated further.

1.6 Life-cycle and Organisational Models

The penultimate chapter brings together much of what has preceded it into a discussion of the practical aspects of managing digital image content over its lifetime, and the repository infrastructure that is needed to support that management. It provides a summary of the OAIS model and the Attributes of a Trusted Digital Repository framework, and attempts to bring these into a simple schematic life-cycle model that identifies the key events that take place during the life-cycle, the activities that are likely to take place at those event points, and the policies and processes which support the activities and the events.

The life-cycle identifies six key events:



At each key event a range of actions are, or should, be taken that will affect the future of the digital images. Many of these actions will affect the longer term survival of the images and will determine if they are merely a collection of bits, or something that remains fit for purpose over time and changing requirements. The key concept is the construction of 'a managed environment' that facilitates the ongoing decision making and actions required to sustain accessibility and usability and to preserve digital objects themselves. The model identifies the key events that take place, the activities that should take place at those events, and the policies and processes which underpin them. Thus the top layer conveys the idea that this is part of continuum where key actions points are identified; the second layer outlines the actions and decisions that are likely to made at this time; and the third layer identifies the requirements, policies and processes on which these decisions are likely to be based.

This section emphasises the concept of fitness for purpose and questions whether we should blindly seek to preserve digital image collections without considering if they are still fit for purpose. It is, of course, essential that we can render and understand the image, but it may be that the quality is such that its value to the user community is so low as to make its continued preservation meaningless. Bringing into the loop user requirements and practices and factoring these into preservation decisions is essential. We suggest that repositories should question whether to continue to preserve the existing image, or whether it may be more useful, and perhaps cost effective, to re-digitise - where this is possible of course - to improve its quality and to take advantage of new technologies and methods for improving the colour space or the resolution. At the present time we have no formal methodology and decision making process built into archival and preservation practice to undertake these kinds of decisions. **We therefore recommend that JISC consider commissioning research into when this type of approach might be appropriate, along with the associated costs. This should form part of a wider research agenda to consider selection and retention policies in the digital age.**

The chapter also discusses a range of organisational models for consideration, ranging from single site models through to a number of different formations of disaggregation. These include:

- Single Institution Image Repositories
- Image Repository with specialist support for preservation
- Image Repository with outsourced preservation services
- Outsourced image repository services

It makes the case that two key factors need to be considered if both fitness for purpose, and the concept of the managed environment are to be addressed. These are that images arise from and are used by subject specialists – visual arts practitioners, architecture students, geographers, and so on. Users of these images will also come with a subject 'eye', and thus subject requirements must be considered for preservation. In addition, there are technical characteristics, many outlined in this report, that define digital images, and these must also be considered when preserving them. We therefore recommend that some form of disaggregated model that can encompass both these needs should be developed. **JISC is recommended to investigate the feasibility and form of possible disaggregated models for the preservation of digital images that can encompass subject, technical and infrastructure requirements.**

1.7 Assessing Costs

Lastly, the report discusses the likely costs associated with the preservation of digital images. It provides a brief overview of the methods for assessing costs but is unable to give figures, largely because we do not yet have the practical experience to do so. However, there are an increasing number of studies and reports on assessing costs, and some repositories have identified the costs of establishing and running a repository. It would be useful if the community had access to these studies and reports in a single place, alongside a commentary on the work undertaken so far. This would enable image repositories at least to define their methodology for assessing likely costs and assist in the development of appropriate business plans that would ensure financial stability and sustainability – essential if a repository is to conform to the TDR requirements. **We therefore recommend that the JISC build on the outcomes of the LIFE project and the ERPANET Cost Orientation Tool to build an on-line tool that would enable repositories to assess likely costs against a range of variables and outcomes, and to review reports and papers discussing cost issues. The DCC might be an appropriate body to take this work forward.**

2. Introduction and Scope of Study

2.1 Outline and Scope of Study

UK Higher and Further Education institutions, alongside many other cultural heritage and other organisations have invested considerable effort and resources into digitising or capturing in digital form raster and vector images for use in research, teaching and learning. To secure the long-term future of these digital assets, significant effort must now be put into ensuring that they are preserved and continue to be accessible in the future.

2.2 JISC activity and initiatives

JISC has played a significant role in advancing the digital preservation agenda both in the UK and internationally:

- Funding a series of seven digital preservation research studies as part of the eLib programme
- Jointly (with the AHRC and the ESRC) funding the Arts and Humanities Data Service (AHDS), and the Economic and Social Data Service (ESDS)
- Funding the Cedars digital preservation project, and the CAMiIEON digital preservation project
- Establishing the JISC Digital Preservation Focus in June 2000 to develop the JISC Preservation Strategy that outlined activities from 2002-2005
- Funding the FAIR Programme, including a preservation stream
- In partnership with other organisations and sectors, establishing a Digital Preservation Coalition aimed at developing the UK digital preservation agenda in an international context
- In partnership with the e-Science Programme, establishing the Digital Curation Centre to help solve the extensive challenges of digital preservation and to provide research, advice and support services to UK institutions
- Funding a number of studies into the preservation of different digital objects, inlcuding e-prints and teaching and learning objects

JISC recognises that the increasing scale and complexity of digital resources now requires a shift in emphasis from relatively modest funding for research into digital preservation towards the establishment and on-going support of shared services and tools. Digital preservation represents a complex set of challenges, which are exceptionally difficult for institutions to address individually. National action in this field is therefore appropriate to the community and remit and mission of the JISC.

JISC's continuing commitment to developing the UK digital preservation agenda was set out in the JISC *Continuing Access and Digital Preservation Strategy 2002-5* (Beagrie, 2002). In this Strategy JISC envisaged responsibility for digital preservation activities spread between national services, individual institutions and, potentially, institutional consortia. The Digital Curation Centre (DCC) arose from the Strategy and will act as a conduit for sharing expertise and developing best practice. The DCC will itself not hold digital resources, but will provide a set of central services, standards and tools for digital repositories.

This study follows on from the numerous activities outlined above. The authors expect it to impact in two key areas:

- First, as a source of knowledge and information available to the creators and curators of both vector and raster digital images
- As a source of information and knowledge for JISC as it develops its priorities for preservation activities over the coming years.

2.3 Outline scope of study and methodology

The study was conducted from August to December 2005 by a team from the Arts and Humanities Data Service (AHDS), including the AHDS Executive, AHDS Visual Arts and AHDS Archaeology. It was conducted in response to the JISC Invitation to Tender for a Digital Image Archiving Study, the purpose of which was to:

"Carry out a scoping exercise on the preservation of digital image files as well as determining archiving methodologies and future research possibilities. The study will include both 'born digital' and digitised material."

The definition of a digital image used in this study is as follows:

"Digital (still) images are non-moving representations of visual information"

That is, still images that convey their meaning in visual terms, e.g. photographs, posters, diagrams, drawings. The study considers both the familiar raster image and the perhaps less well known vector image. The former include the products of digital photography and scanning with file formats such as TIFF and JPEG. The latter, although it may be considered less when thinking of digital images, is probably more common as the almost ubiquitous PDF file, a subset of PostScript, is a vector-based format. Both can be said to be geometric or spatial but any similarities end there.

Raster (or bitmapped) images are grid-based with information being held about each point or pixel within the grid, whereas vector images have information about any number X, Y, Z spatially defined coordinates and are made up of scalable objects—lines, curves, and shapes—defined in mathematical terms, often with typographic insertions.

2.4 Project Outcomes

This report presents the findings of the study, arranged into four main areas:

- User requirements
- Technical Characteristics of Digital Images
- Metadata and Preservation Methodology for Digital Images
- Modelling the Life-cycle for the Preservation of Digital Images

The study focuses on the requirements for the *long-term* preservation of digital images, which is defined, for the purposes of this report, as the period of time during which the hardware, software, and standards used to create and access digital objects become obsolete and can no longer be obtained. Digital objects store meaningful information encoded as a stream of binary digits (bits). In addition to preserving the bit stream (bit preservation), and ensuring that it is not destroyed or corrupted, digital preservation involves ensuring that the bit stream can be correctly decoded and converted into meaningful information again (functional preservation). This report considers both these aspects of digital preservation.

However, the study also brings a user driven approach, seeking to identify user requirements and the role these must play when developing preservation strategies and implementations. Digital preservation involves active intervention across the life-cycle of a digital object. The long-term survival of digital objects will be affected by the priorities and actions of all those who have an interest in it. There are three main stakeholder groups with an interest in digital images: creators, users, and repository managers. The study team sought to contact and receive feedback from a wide range of stakeholders, including those creating image content, those using image content and those responsible for managing and preserving image content. We recognise that many of the stakeholders contributing to the study combine one or more of these functions.

The recommendations made in this report will primarily affect repository managers, but in some cases are also highly relevant to both creators and users. It is the belief of the authors of this report that all those with an interest in using digital content, whether in the short-term or the long-term should be concerned about ensuring continued access to digital content – and by default, its preservation.

Even when recommendations are targeted at those who fund, plan and manage repositories, the importance of ensuring appropriate involvement from creators and users of digital images should not be forgotten. Ultimately, the value of digital images is in their value to users, both now and in the future. Repository managers must meet the needs of long-term preservation in a way that does not conflict with the requirements of the user. Any preservation strategy must keep user requirements in mind and be responsive to actual needs in the present, and changing needs over time. The logical conclusion to draw from this approach is that any preservation strategy and actions must be flexible. Preservation is not a one-time activity – it is an ongoing and changing process where policies and practices will need to be adapted and change.

Comments on this report may be directed to:

Sheila Anderson
Director
Arts and Humanities Data Service
King's College London
26-29 Drury Lane
London, WC2B 5RL
Email: sheila.anderson@ahds.ac.uk
URL: http://ahds.ac.uk

3. Context and Background

3.1 Overview of Current Research

Librarians and archivists have studied digital preservation issues for the last twenty years, but no consensus has been reached regarding a shared digital preservation strategy.[3] A number of digital preservation strategies have been proposed, but there is no definitive approach to the problem of maintaining digital content across multiple generations of technology.

Management and preservation requirements for digital materials are fundamentally different from analogue materials. Digital materials can be created using a wide range of technologies and formats, whether born digital or digital surrogates of existing analogue materials. They can be described and documented in a variety of ways – or not at all. They are subject to both physical deterioration and technical obsolescence. More than one copy can be easily and simply created. Access may be provided through more than one point, and may be distributed. All these factors will impinge upon the approach taken to their management and long-term preservation.

These differences present the curators of digital materials with some fundamental challenges. The way in which materials are created, particularly the technologies used, will determine how conducive to long-term preservation the materials are, and will present varied challenges to curators charged with the subsequent management and preservation of the materials. Curators will need adequate metadata about the resource if they are to successfully manage, preserve and make the materials accessible. Multiple copies may also imply multiple versions the digital resource curator must somehow ensure the integrity and authenticity of the resource. They must be aware of changing technologies and fragility of media and take these into consideration from an early stage in the ingest process.

Within the digital preservation community, the concept of the life-cycle (or continuum as it is sometimes called) management of digital resources has emerged to describe and document the active management processes that need to take place, and the key decision-making and intervention points along the continuum. The life-cycle concept has been incorporated into OAIS Reference Model, now adopted as an ISO standard for digital preservation. The OAIS model is proving a strong foundation for the development of digital archiving projects and services, and is increasingly being implemented by digital libraries and archives, including the AHDS.

Some studies suggest that a national preservation policy is a necessity,[4] whilst others embrace diversity in preservation policies. A recent article, 'Eternal Bits: How Can We Preserve Digital Files and Save Our Collective Memory?' states that 'There is no right way to preserve digital content. Just as biodiversity is good for the natural environment, different digital preservation policies and strategies are good for the preservation environment'.[5] A number of digital preservation strategies have been proposed, but there is no definitive approach to the problem of maintaining digital

---

[3] Seamus Ross, 'Changing Trains at Wigan: Digital Preservation and the Future of Scholarship', NPO Preservation Guidance Occasional Papers', (London, National Preservation Office, 2000) p. 13.

[4] See for instance Barbara Bültman, Rachel Hardy, Adrienne Muir and Clara Wictor, *Digitised Content in the UK Research Library and Archives Sector: A Report to the Consortium of Research Libraries and the JISC,* (London, 2005).

[5] Mackenzie Smith, 'Eternal Bits: How Can We Preserve Digital Files and Save Our Collective Memory'? *IEEE Spectrum,* (July 2005).

content across multiple generations of technology. Unfortunately, information on the likely costs, possible limitations and long-term sustainability of different strategies is far from complete – partly it must be said, for the very valid reason that no one has yet the time to gain the experience needed to answer these questions.

Current and upcoming research is flagged up through the Digital Preservation Coalition's quarterly, 'What's new in digital preservation'. The DPC have launched the UK Digital Preservation Needs Assessment Survey, to reveal the extent of the risk of loss or degradation to digital material held in the UK's public and private sectors. 'Preservation Strategies' are to be featured in the DCC Digital Curation Manual (**http://www.dcc.ac.uk/resource/curation-manual/**). This instalment will examine some of the technical issues surrounding digital preservation and also explore some of the philosophical issues that may hinder effective uptake and implementation of digital preservation strategies. Research Libraries Group DigiNews have reviewed digital preservation and digitisation research and development projects around the world and have compiled a list of 10 initiatives in progress that have the potential to contribute to the broader cultural heritage community. These include:

- The National Digital Information Infrastructure and Preservation Program (NDIIPP) - ECHO Depository Project
- The Digital Archiving and Long-Term Preservation (DIGARCH) program
- Supporting Digital Preservation and Asset Management in Institutions program of the Joint Information Systems Committee (JISC) project DAAT: Digital Asset Assessment Tool
- Metadata Generation Research (MGR), 2005-
- The SURF funded Digital Academic Repositories (DARE), 2003-2006
- Australian Partnership for Sustainable Repositories (APSR), 2004-2006
- Digital Curation Centre (DCC), 2004-2006
- The Andrew W. Mellon Foundation supported Auditing and Certification of Digital Archives, 2005-2006
- DELOS Digital Preservation Cluster, 2004-2006
- The NARA Virtual Archives Laboratory (VAL)

A number of these projects are examined in greater detail in Appendix A.

Nevertheless, despite these developments, the difficulty for those undertaking preservation or with the responsibility for providing access in the long term or digital is the lack of practical advice, and of robust tools and mature techniques for digital preservation. Research tends to be theoretically-based. Testbed projects have begun to emerge, but overall research tends to be general in nature. Work specific to digital image preservation is rare. A select list of support and advisory bodies with preservation responsibilities, JISC activity, research and practical research projects are available in Appendix A. At this stage the number of projects underway have not reached the critical level required to push the preservation agenda forward. It can be assumed that continued practical experimentation, conferences and awards will advance the preservation agenda. This expectation is a feature of recent preservation studies.

4. User Requirements

4.1 Introduction

This chapter examines the considerations that must be given, in the digital image preservation process, to the users of those images. In light of Neil Beagrie's identification, in 2000, of a need to establish strategies and services for long-term *access* of digital resources in UK HE/FE communities,[6] the section considers how access is driven by user needs.

It is now widely understood that the development of digital systems, like any non-digital system, should be led by an understanding of the needs of those people who will use the system. The JISC-funded Digital Repositories Review states that "it is vital that repositories meet the needs of users; there is a need to explore user requirements and prioritise them in the development of repositories; the process needs to engage the user community in a real way."[7] On a bigger scale, the House of Commons' Education & Skills Committee report following the low uptake of the £50m UK e-University noted that it "failed largely because it took a supply-driven rather than demand-led approach"[8]. Consequently, this section explores the impact that taking a user-led perspective has on the development and implementation of digital preservation processes. Obviously, the needs of *end*-users are of primary importance, i.e. those groups of people who will require access to the preserved images at some time in the future. The preservation process must necessarily meet not only the end-user needs integral to the image object itself (e.g. file size, colour attributes), but also those relating to accessibility, integrity and authentication. Furthermore, it is equally important for the process to make consideration to the *original* users, or creators of images, particularly since they make decisions that will, inevitably, impact on the success or failure of any onward preservation and access of those images; for example, decisions about quality, metadata or usage rights. These are discussed here in terms a user requirements analysis, but more detail regarding explicitly technical issues for creators, is given in the State of Art Review section of this paper.

4.2 Identifying users and potential usage of digital images

In order to deal with issues for users of digital images, we must first define what we mean by users and, subsequently, identify just who the actual people involved might be.

Broadly speaking, users of digital images can be divided into two categories: those who create digital images or are involved in their creation through commissioning, managing or perhaps financing the creation; and those who make use of those images as part of their work through, for example, manipulation of the image, study of the subject matter or interest in the digital object itself. Although any system or resource should be developed in light of its end-user (it's raison d'être), imperatives from 'behind the scenes' can often be more influential; for example, the requirements of legal entities, branding experts or funding bodies. In terms of a preservation process, it is therefore necessary to explore the needs of both sets of users, and to establish the impact that each will have on the other's needs or requirements. Both

[6] Neil Beagrie, JISC Digital Preservation Focus and the Digital Preservation Coalition.
[7] S. Anderson and R. Heery, 2005, 'Digital Repositories Review' (2005), p. 15.
[8] House of Commons Education & Skills Committee, 'UK e-University': 3rd Report (2004-2005), p. 3.

groups are 'stakeholders' in the strategies, processes and systems needed for preserving digital images.

Perhaps, for ease of reference, the behind-the-scenes participants should be referred to as 'Start-users': the individuals or organisations who contribute to the development, deployment or expansion of the resource to be preserved, or to the preservation system itself. Figure 1 suggests some examples of the differing *end* and *start* users that may have to considered in the development and/or implementation of an image preservation process, though it is important to recognise that there will always be overlap between the two groups (some users will wear different hats at different times).

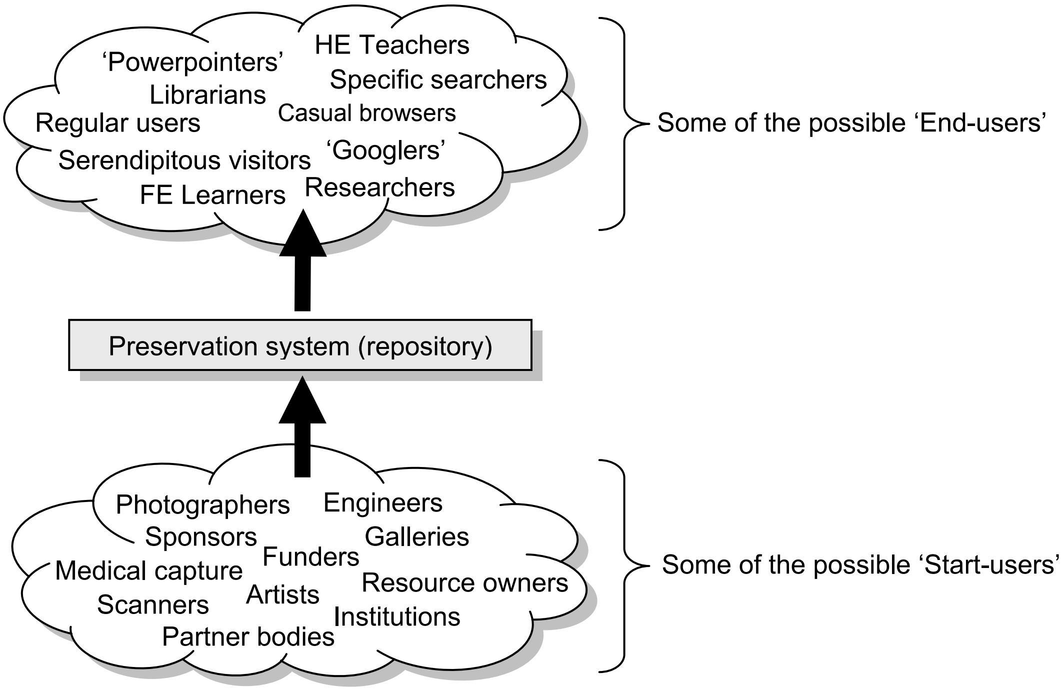


Figure 1. Examples of stakeholders on both sides of the digital image preservation process.

4.3 Who are they?

As the figure above demonstrates, users of digital images can be defined as two simple groups. However, the figure also illustrates just how varied the participants in those two groups can be.

There are a number of ways that user groups can be identified to provide a cross section of the UK HE community as relevant to the JISC. They can be considered in respect to both larger scale environments and to smaller, more specific areas of usage. A comprehensive list of content creators that affect the JISC community can be found in State of Art Review, but some of the broad categories are presented here as the basis for subsequent analysis.

The users whose needs must be considered are described here in terms of the HE education sector as well as particular aspects of the collateral environment; for example, users of museums/galleries or staff/students of Further Education institutions. However, usage of images will clearly cover a broad spectrum of different approaches, and it is necessary to think of a ‘user’ in a number of diverse ways. The following list breaks down the potential users of images in UK education according to several broad and distinct perspectives.

Users of online, education-led services:

- JISC-funded collections (e.g. BioMed, AHDS Visual Arts, Education Images Online);
- International collections (e.g. MOMA - New York, PictureAustralia, the Louvre);
- National institutions (e.g. National Gallery, Victoria & Albert Museum, English Heritage);
- Regional institutions (e.g. Glasgow Museums, Manchester Art Gallery);
- Commercial/subscription-based providers (e.g. Bridgeman Art Library, Wellcome Trust);
- Higher Education institute external providers (e.g. Birkbeck, Courtauld Institute).

Users of internal institution provision:

- Intranets
- Institutional repositories
- Virtual Learning Environments (e.g. Blackboard, WebCT)
- Shared drives
- Digital slide libraries

Users of images in different areas of education:

- Non-vocational;
- Lifelong learning;
- Further education;
- Higher education;
- Masters study;
- PhD study;
- Post doctoral research.

Users with different roles:

- Students;
- Researchers;
- Support staff;
- Practitioner;
- Lecturers;
- Managers;
- Librarians.

Users with different learning needs:

- Dyslexia
- Colour blindness
- Physical impairment
- English as a second language

Users with different subject needs (*based on HERO subject oriented directories*[9]):

- Creative arts and sport
- Humanities
- Engineering, applied sciences and technology
- Languages, linguistics and literature
- Library and information technology
- Philosophy and Psychology
- Religion and Theology
- Social Sciences, Law and Government

Three other, more general, ways of breaking down user groups will also need consideration:

- Users of born digital vs. Users of digitised.
  Different preservation criteria may be important to users depending on whether the image is a digitised surrogate of an analogue object, or an object originally created in a computer environment;

- Users of different generations.
  Diverse age groups may consider different things to be important, in terms of what a preservation process provides. Rather than make assumptions about different age groups, this section will refer only to differences between the needs of 'digital natives' – those who have grown up in the digital environment, and 'digital immigrants' – those for whom the digital environment has been a relatively recent introduction.[10]

- Users of different types of images.
  As this paper has expressed elsewhere, images themselves come in different formats; for example, bitmap and vector. It is likely that the users of differing image types or formats will have differing views on the inherent preservation needs of each.

Considering the breadth of possible use-case scenarios that the above list could produce is, at the least, daunting and, at best, impractical. Particularly given that for every item on the list there will be an equally complex list of further possible scenarios; for example, a category comprising creative arts and sport clearly represents an extremely wide range of different people with a correspondingly wide range of circumstances for using digital images. Furthermore, it is true to say that user needs could be defined in a wider way; for example, the particular needs that an institution or a VLE might have. Nonetheless, for pragmatic reasons this paper will identify needs according to the *people* who actually define those needs.

In order to deal with the potential complexity of trying to establish *all* user needs in the JISC environment relating to the preservation of digital images, four approaches have been taken:

1. **Targeted survey** with representatives from a cross-section sample of the above groups with an emphasis on talking to users with some level of understanding of digital image preservation needs for their particular area;

---

9 http://www.hero.ac.uk/uk/reference_and_subject_resources/subject_oriented_directories3810.cfm
10 Marc Prensky, 'Digital Natives, Digital Immigrants', in *On the Horizon* (Vol. 9 No. 5, October 2001) http://www.marcprensky.com/writing/default.asp.

2. A non-prescriptive **broad survey** was made available through AHDS websites and via email lists, to elicit unprompted preservation needs and specialist requirements.;
3. Creation of profiles of **archetypal users** covering a range of different perspectives and levels of understanding regarding digital image preservation;
4. Desk-top **research** examining previous work on digital image preservation user needs (references are given throughout this paper).

4.4 Targeted survey

To gain a detailed view of a cross-section of the JISC community's perspectives on digital image preservation, a questionnaire was created to enable a number of targeted surveys to take place. The research deliberately targeted people who should, by way of their position, have some understanding of the use of digital images and, hopefully, some understanding of the needs and possible issues surrounding their preservation. This approach ensured that some level of expert user was consulted from across the main areas mentioned in the list in section 2.1.

The questionnaire was initially sketched out through expertise and experience within the AHDS and then augmented by reference to the JISC-funded, Rights and Rewards survey on use of repositories[11], and a questionnaire used to elicit image usage at the University of Newcastle at Northumbria[12]. The questionnaire and results can be seen in Section A1 of Appendix A.

The structured questionnaire was used with seven expert users and informal discussions were had with a further four people.

Summary of results (Most important issues highlighted)

a.	**Importance:**	**b.**	**Main purpose of images:**
	Essential (3)		Lectures (5)
	Important (2)		Publication (5)
	Not too important (2)		Research (4)
			Handouts (4)
			Projects (4)
			VLE (2)
			Exhibitions (2)
c.	**Main problems accessing images:**	**d.**	**Where to access images:**
	Concerns about copyright (5)		Online – WWW (5)
	Lack of metadata (2)		Repository (3)
	Lack of subject relevance (2)		Intranet (3)
	Don't know where to look (2)		VLE (2)
			Own PC (2)
			College server (2)
e.	**Main issues over time:**	**f.**	**Most important aspects:**
	Copyright (5)		Subject (5)
	Changes in technology (3)		Flexibility of use (3)
	Quality (2)		Pixel quality (2)

11 http://rightsandrewards.lboro.ac.uk
12 Margaret Graham, 'Use and impact of digital images for teaching, learning and research in Northumbria'.

	Ability to search (2)		
	Provenance (2)		
g.	**Caring for images:**	**h.**	**Access control:**
	One National body (5)		Free access to all (3)
	National subject-based bodies (3)		Registered users only (2)
			Different levels of access (2)
i.	**Quality control**		
	Agreed quality standards (4)		
	Metadata schemas (3)		
	Meeting technical requirements (3)		
	Legal checks (2)		

4.5 Broad survey

To ensure that a broader view of user needs relating to digital image preservation was elicited, a survey was conducted through AHDS websites and via email lists. The survey included a very short, non-prescriptive survey (consisting of just three questions) to permit users to provide whatever information that they deemed appropriate or important. The survey elicited 101 responses between Tues 25th Oct and Monday 14th November 2005. Results can be seen in Section A2 of Appendix A.

a. Main issues raised:

Metadata:	54
Access/usability:	52
Copyright:	45
Quality:	44
Formats/media:	32
Changing platforms:	14
Preservation/back-up:	12
Prescriptive guidelines:	10
Costs:	9
Retrieval:	8
Responsibility:	4
Selection:	4
Strategies	3

b. Further issues raised:

Indexing suitability of images for education (2)
Understanding the science behind preservation (2)
Developing descriptive system for image content (2)
Involving everyone, not just slide librarians (1)
Sensitivity of data (1)
Encryption (1)
Central repositories (1)
Publicity (1)

c. Main areas of respondent interest:

Art History:	17

Digital Asset Management:	10
Archaeology:	8
Performing arts:	8
Art & Design:	8
History:	7
Photography:	7
Museums:	6
Lit/Lang/Ling:	5
Information Technology	4
Libraries:	4
Conservation:	4
Arial photography:	3
Bookbinding:	2
Imaging:	2
Cultural heritage:	2
Film:	2
Data development:	2
Arts & Humanities:	1
Social Sciences:	1
Philosophy:	1
Medicine:	1
Classics:	1
Crime/Forensic med:	1

4.6 Taxonomy of 'user's scenario's'

4.6.i Archetypal users

Below is given a cross-section of differing, archetypal users; each created from combinations of elements outlined of the previous sections. Having such archetypal models, added to the three further modes of user needs identification, helps to place in context the most significant attributes of digital image preservation, that are important to, or affected by, users.

David is a fifty two year old archaeological illustrator. He is a part-time lecturer at an art college where he teaches HND, BA and MA students, and he spends the rest of his time as a freelance illustrator in the heritage sector. As a digital immigrant, David has observed that the work he and his students now create using digital technologies is not afforded the same protection for the future as the work that he used to do on paper. Perhaps because of his work in the heritage sector, David is well aware of the need to use the right, acid-free papers and the idea of keeping items, well catalogued, in suitable physical environments so that they are available for future generations to access. He is worried that much of what is now being created is being lost at an alarming rate and he sees some irony in the fact, for much of his digital work, the only copy that is being properly preserved is the final printed paper version.

Alan is a twenty four year old graphics lecturer. He has recently taken up a post teaching a number of modules, mostly to BA level students, in advanced graphics techniques ranging from three-dimensional modelling to video editing. Alan loves the power and scope of the new digital technologies and is a keen ‘gadgets’ man, owning the latest mobile phone and palmtop device. He has been slightly taken aback to find that many of the things he learned on his own degree course are not available in his new post. In particular he cannot understand how that college looks after any of the student’s work given that it does not have any sort of repository services. Such is his enthusiasm, Alan has proposed to his head of department that the graphics department invests in its own repository server, rather than wait for an unknown length of time for the college to provide a suitable service. He is confident that he knows what is needed and has already provide an outline specification that he thinks will meet the needs of his students.

Mary is a thirty six year old librarian. She spends much of her time at the ‘front-line’, answering queries and assisting students and staff at a busy library help-desk in a college that caters for both FE and HE study. Mary is quite used to archiving material and has a thorough understanding of metadata and copyright issues. Lately, Mary has become worried about the increasing pressure to meet what she sees as ‘impossible demands’ on her time and the library’s resources to reach a number of her institute’s strategic aims relating to digital usage. In particular she is concerned that all manner of material is being created and placed on the intranet, VLE and internal servers with no regard to copyright or quality. She is also bothered that, because no-one is recording any of the material creation properly, no-one else will be able to access it in the future, especially after the creator has moved on to new pastures.

Simon is a forty four year old scientist. He lectures BA and MA students in systems engineering and formal specification at a university that specialises in scientific subjects. Simon is something of a digital immigrant but, having worked with computers throughout his adult life, is extremely comfortable with the technology and with the rapid pace at which it evolves. Nonetheless, Simon is equally aware that, with every new development, there is the risk that previous work can vanish. His usual take on this is that, so long as work has been done scientifically and recorded and published properly, then the inherent information will always be available. However, Simon has also seen how much of what happens in the digital environment of his institution can be affected by departmental and political changes. As he sees it, most of the day-to-day work is conducted with a view based on the needs and imperatives (financial, political or educational) of today, rather than tomorrow. He fears that, as such, much of the material on the College intranet or servers will not necessarily be ‘migrated’ every time there is a change in circumstances.

John is a twenty three year old waiter. Since leaving school at sixteen, John has had a ‘chequered career’ that has finally brought him to the conclusion that he wants to return to education. Consequently, he has enrolled on a part-time class in HNC Computer Aided Design in the hope that he will be able to move into kitchen design. Much of his time is spent working with CAD software and, although such things are presented during his course, he has little interest in issues such as copyright or preservation. Nonetheless, when his coursework involves research into kitchen design and the creation of essays or reports, John does wish that there was some easy-to-use, comprehensive source of design material available, especially in regard to images of kitchens. Most of the time he just uses Google, which usually provides him with the images that he needs, even if they are not very good quality.

Silvia is a thirty six year old researcher. She has a PhD in Medical Information Management and runs a small, successful University research team specialising in clinical recordings deposit and access for HE students and researchers. Silvia fully understands the needs for preserving digital images and, in her role as medical adviser to the University IT department, has been involved in the University's setting up of a first class institutional repository. However, as a researcher she worries about the increasingly complex nature of laws and rights issues within the medical world and the ever rising number of IPR and copyright dilemmas within her institution. Silvia is beginning to feel that there will never be an answer to providing broad or long-term access to clinical recordings, and is frustrated at the burden of complicated strategies and policies around the issues, whilst all the time she struggles to ensure access to the images that her staff and researchers need.

4.7 The users' perspective

The four approaches to identifying user needs have been collated, and an outline of responses, thoughts and ideas is assimilated in the following subsections:

4.7.i Educational needs

Most users of images in education are primarily interested in their preservation as it relates to teaching, learning and research aims and practices. Many of them, particularly lecturers, use images for presentations and, traditionally, the slide projector has been the main tool of this trade. However, as many digital immigrants are now discovering, there has recently been a significant shift from the traditional approach, exemplified by the fact that well-known slide projector manufacturer, Kodak, announced that production and sales of their machines would end in June 2004, with service and support ceasing in 2011. In September 2003, the company issued a press release explaining their decision in light of the fact that: "..in recent years, slide projectors have declined in usage, replaced by alternative projection technologies."[13] Although, most users are quickly converting to the new digital display technologies, such as PowerPoint, the dramatic and in some cases unwelcome transition, is a cause of concern when considering preservation issues: users worry that sudden developments in new technology could still render all the current work to preserve digital images obsolete in the not too distant future.

The rapid change in technological approach also stimulates another concern relating to the educational needs behind the preservation of images. This is the notion that pedagogical perspectives have always been subject to change, in response to stimulators such as government edicts, new research and altering social fashions. Consequently, it is possible that any process for preserving digital images today, no matter how technologically stable, will not necessarily be capable of evolving to meet changes in educational needs of the future.

One possible implication here is that traditional user views of archives and preservation processes may have to change fundamentally, to a model which can itself adapt to both technological and pedagogical changes, 'on the fly'. In essence, the management processes and expectations in the user domain may have to 'migrate' just as the digital object must – a switch from "batch processing" to "continuous processing" [14]. Whilst such associated concepts as "depreciation" and "lifecycle costing" may be familiar elsewhere, they are new ideas to a sector where resources are traditionally built up over long periods of time and preserved on the shelf, with relatively minimal maintenance, for future students and staff. With such

13 http://www.kodak.com/US/en/corp/pressReleases/pr20030926-01.shtml
14 M.S. Lynn, 'The Impact of Digital Technologies in The Relationship Between Digital and Other Media Conversion Processes: A Structured Glossary of Technical Terms', http://palimpsest.stanford.edu/byauth/lynn/glossary/impact.html.

notions comes added complexity, and need for more training and human resources and, inevitably, an increase (or significant change) in costing models.

The educational needs relating to the digital object itself, such as selection procedures and quality issues, are described further on, in section 3.4.

4.7.ii Subject distinctions

Continuing on from the previous section, we can see that, as overall educational practices may have to adapt to make full use of new technologies, different subject areas will have different perspectives on such adaptation. Consequently, representatives of different subject areas may perceive different values in the process of preserving, and then accessing, digital images. The range covers everything from, as one expert respondent commented: “A broad range, some as pure eye candy to those that are intrinsic to the learning episodes”.

However, trying to identify such differences proves to be a difficult task. It seems that many apparent difference are simply varying perceptions based on principles, processes and terminology within different disciplines, and they do not directly impact the way an image should be preserved. Most of the different requirements can be met, if digital image preservation processes work to the highest possible technical standards and with a view to preserving the main attributes of an image, as defined below. In essence, there is a lowest common denominator for digital image preservation.

4.8 Dissemination/access preferences

There is only one reason to preserve digital images: so that they are available to people in the future. As the results of the broad survey indicate, to most users of digital images access and metadata are the two top areas of concern regarding preservation. The following quotes from the survey illustrate this:

“As the whole point of preservation is the preservation of access, continued access is clearly the most important issue.”

“There will always be preservation issues but access is key.”

“People seem to work from the point of view of the digi-expert, web designer, archive owner. The user should be considered first.”

“Access (that's the point, isn't it?).”

And, regarding the inextricable link between access and metadata:

“Access: what's the point of *the image* existing if it can not be accessed? Metadata: there is no point in *the image* existing if it can not be found.”

As is the case with any knowledge source, digital or otherwise, appropriate indexing or cataloguing of the resources is imperative. In a digital information system, whether it be for short or long term access, such ‘descriptions’ of the object are held in metadata fields. This is particularly relevant for the archiving of, and subsequent access to images, since by virtue of their visual nature, they commonly travel without any intrinsic (textual) information. Without understandable metadata, there is significant risk that the future will produce millions of images that will be unsearchable, unidentifiable and unusable.

More detailed information regarding metadata is included elsewhere in this paper, but from a user's perspective, there are some general issues that must be considered in the preservation of digital images. These mostly relate to the difficulties of preserving terminology, with the images, that will ensure their usability and re-usability across a vast range of potential users, in an infinite number of ways and into an unknown future. Two areas of terminology need to be considered:

i. **Content-specific terminology.** To enable different users to successfully access material in the future, different considerations need to be given to differing approaches. Usually this involves creating metadata for particular images based on the subject area within which that image sits. This is achieved through the use of specific metadata schemas. As mentioned in 3.1, however, user practices change, and in turn, so can the terminology employed to find images. If the user terminology alters over time then the usefulness of the preserved images, with their terminology based on a fixed point, will diminish. Lehman's Laws (relating to software evolution) all point to the need to accept change as intrinsic to the development of any system. In particular, his first law states, "in accord with universal experience, that a system must be continually adapted else it becomes progressively less satisfactory" [15]. This is as true of the metadata and other descriptive terminology employed as it is of the objects or the system itself.

ii. **Technology-driven terminology.** As mentioned in section 3.1 most users are concerned with achieving educational aims and, consequently, have little need (or desire) to understand *mega-pixels, bitstreams or emulation techniques*. As the digital preservation project at the British Library found out, it is important to users that jargon[16] is kept to the necessary only. Particularly when designing systems and processes that are to be understood in the future, any terms used need to be either explained as part of the process (e.g. through glossaries or thesauri) and/or, where possible, presented in as plain a language as possible. This is also important because of the fact that different people will read different meanings into such terms; for example, in the Cedars project, it was noted that, "the relative novelty of digital preservation as an issue for libraries, and the fact that expertise in this area resides in other sectors (e.g. electronic records management) means that defining what we mean by specific terms is sometimes contentious. Where librarians, archivists, records managers and computing technologists assemble, the term "archive" can (and does) mean very different things."[17]

Both of the above issues indicate a need for acceptance that terminology within a preservation system needs to be considered extremely carefully, else it may reduce the success with which users will be able to access the material within.

However, in terms of access to *all* users, preservation systems need to also consider the important issues of accessibility for disadvantaged users. Images pose a particular problem for some users, particularly, for example, those with visual impairments, and it is essential that a system for providing long term access is fully

15 http://www.doc.ic.ac.uk/~mml/feast2/papers/pdf/556.pdf
16 Helen Shenton, 'From Talking to Doing: Digital Preservation at the British Library', (The British Library), http://www.rlg.org/en/page.php?Page_ID=255.
17 'CEDARS: Long-term Access and Usability of Digital Resources: The Digital Preservation Conundrum' http://www.ariadne.ac.uk/issue18/cedars/

compliant with guidelines for accessibility (e.g. W3C[18]) both now and in the future. It may be that the standards, approaches and technologies of the future permit greater accessibility to digital images for a wider range of users, and a preservation system must be equipped to accommodate such change.

4.9 Impact of user requirements on preservation methods and approaches

The following sub-sections break down the areas that seem to be of most concern to most users in respect to the preservation of digital images.

4.9.i Image quality

The quality of images is generally seen as an important issue for most users. However, the definition of quality itself is open to a number of different interpretations when discussing the attributes most significant in a preservation process. Nonetheless, most quality issues can be described in one of two ways:

- Whether the **content** is sufficient for the task. This qualitative, and usually subjective, value is based on things such as the subject matter itself: whether the digital image offers a good representation of its subject; and whether it fits in with the views, tastes and current needs of the user. Such issues have long been applicable to pictorial representations and do not seem to alter significantly when addressed to digital images.

- Whether the **object** itself is good enough for the task. Although this issue can also be subjective, its most significant elements relate to more technical attributes; for example, the level of detail (LoD), colour depth, pixilation or resolution, and are based on an assumption that the content is satisfactory

In both cases, the overriding factor in user decisions about quality is *fitness for purpose*. Different users will have very different ideas about what constitutes quality in either sense

In terms of the quality of the object, again different users can have very different definitions of fitness for their purpose; for example, a history tutor talking about Winston Churchill to an FE class may only require a 72dpi copy of a grainy, black and white photograph to illustrate his/her talk, whereas a geology researcher trying to identify features on aerial photographs may require very high definition, high resolution images to be sure of his/her findings.

Unfortunately, one issue that cannot be overcome by this approach is the fact that, as display mechanisms improve, today's very high quality image will, by tomorrow's standards, be of an increasingly poorer quality. There is a widespread belief among many user communities that quality of digital images can be defined in terms of numbers of pixels; an image that is 2000 pixels across and 300dpi, is considered a 'good quality' image. What is less understood is that, for preservation purposes, this is only a relative measurement. In ten years time, such an image will probably be considered 'poor quality', particularly as display mechanisms improve. The solution to this dilemma, in defining possible user needs of future communities, may lie in

[18] http://www.w3.org/

Nicholas Negroponte's observations on the early development of digital television screens. He emphasised that fact that scalability is the key to ensuring digital growth and that "when people argue over the number of scan lines, the frame rate, or the aspect ratio of television in the future, one can rest assured they are discussing the most irrelevant pieces of the puzzle." He follows up this assertion by stating that the future of all digital media will not be to do with these sorts of erroneous 'standards': "it will be nothing but a bit stream."[19] As mentioned above, the quality of an image should relate to its fitness of purpose; no-one judges the quality of a television programme according to the number of lines on the screen.

However, amongst the concerns there is also some confidence in the education community. One respondent of the broad survey commented:

> "Preservation - understanding the nature of an image that doesn't actually exist in a hardcopy format, e.g. only a bunch of numbers and the ability to keep this image through time on various media, which have yet to be tested. I'm not too worried about not being able to access them as all it takes is the ability to write a program to interpret those numbers on whatever machine, with whatever display. I don't feel these images will ever be 'lost', just might be difficult to write a program to access them. I don't think the future will be as scary as everyone thinks its going to be, so I'm not too worried about images lasting just like paper documents have for the last so many centuries."

Nonetheless, many users do worry about just what is being preserved, though most would probably agree that, in terms of preservation processes, it is safe to assume that preserving the highest available quality of object, in as many respects as is possible, is the best way to serve the future needs of users of digital images. The following sub-sections help to determine exactly what constitutes a 'good quality' image.

4.9.ii Selection procedure

Arguably more important than 'pixel quality' to many users of images, is ensuring that the right images are chosen for preservation in the first place. Sadly, not all digital images that are created today will be saved for tomorrow, and, although there are arguments for a cover-all approach, there are a number of advantages to the eventual users in a more critical selection procedure. The National Library of Australia, in its own plans to digitise material, point out that, not only has this been the normal case for non-digital material, but, selecting a limited number of items to be preserved enables creators to devote more time to: inject quality control into the collecting process; negotiate explicit access rights; develop relationships with creators and publishers; maintain working knowledge of changing technological features and file formats; realistically record metadata for future preservation decisions; and, realistically commit to maintaining access to the material. [20]

Nonetheless, such an approach does provoke questions about just how such a process will take place. For example, it is currently very difficult to maintain the integral structure of a website in any way that is guaranteed to survive into the future.

---

[19] As Scalable as the US Constitution, Nicholas Negroponte, http://web.media.mit.edu/~nicholas/Wired/WIRED1-01.html

[20] Colin Webb, 'Towards a Preserved National Collection of Selected Australian Digital Publications', in 'Preservation: An International Conference on the Preservation and Long Term Accessibility of Digital Materials' (National Library of Australia, 2000), http://www.rlg.org/en/page.php?Page_ID=247.

This means that, simply because of 'technical' impediments, many potential valuable websites will vanish. This sort of selection is, of course, far from ideal for the integrity of preserving educational material for future generations. Beyond that, users worry that decisions will be made on the grounds of lack of perceived quality (in pixel terms), or lack of support at departmental, institutional or National level because of political or funding criteria rather than understanding of educational need.

TASI, the Technical Advisory Service for Images, in discussing a digitisation project, suggest that a 'selection criteria' list is drawn up. Their own example recommends that "the following should be considered:

- Your users - who is likely to use the collection, how will they use it, and what will they expect?
- The uniqueness of your collection - are any of these images available, or likely to become available, elsewhere? Are there a lot of duplicates within the collection?
- The copyright situation - do you have the right or permission to digitise this material? If not, how easy will it be to obtain?
- The form and condition of the material - will your originals require conservation work before they can be captured?
- The metadata - is there enough information about the images to ensure easy retrieval from a database?
- The technical issues - how easy is it going to be to capture these images?"[21]

Beyond this, however, is the need to determine exactly what is captured and preserved within the image, particularly when talking about digital surrogates of analogue images. The report of the US Library of Congress's Manuscript Digitisation Demonstration Project in 1998, noted that perfect reproduction was not necessarily the be all and end all. Rather, it is a single consideration in a strategy of what should be captured and, for their own uses (i.e. scanned versions of manuscripts), safety of the originals, efficiency of production and cost were equally important considerations.[22] The report also noted that, in the goal for higher levels of capture, more human intervention would be required – this would result in slower production time ergo less images captured, and, of course, increased costs and productivity overheads.

4.9.iii Provenance

For end-users of images in an educational environment, one of the most important factors is being able to establish the reliability, authenticity and authentication of the source from which a reference is taken, whether that be a textual reference or an image. This is increasingly difficult in the free-flowing hypertext world and users believe that mechanisms and processes for attributing provenance to digital objects is an extremely important factor in their future usability. Across the broad scope of the UK education community, there are inevitably many ways that such provenance can be defined. But a Nancy Brodie cites a number of different views surrounding the authentication issue; for example "definition and preservation of those features of an information object that distinguish it as a whole and singular work", or "Authentication provides verification that a digital object is what it purports to be and contains the contents that the author/creator or publisher originally intended." She sums up by saying that most definitions "have in common an understanding that authentication is

[21] http://tasi.ac.uk/advice/creating/selection.html

[22] Library of Congress, 'Manuscript Digitization Demonstration Project, Final Report', (October 1998), http://memory.loc.gov/ammem/pictel/

a process which allows future users to determine the authenticity of objects or proves the authenticity of objects", and that there will be different levels of authenticity for different communities as well as different levels of authentication.[23]

Currently, much of the UK education sector uses Athens as an authentication firewall, and most users tend to assume that, if an information source lies behind it, it is probably of a reasonably safe form to be used in an educational context. But, as more and more digital images become available, perhaps via a VLE or institutional repository, such confidence in the integrity and authenticity of the object is diminishing. Users are concerned that, if images are to be preserved, how will their integrity or authenticity, their provenance, be preserved with them.

4.9.iv Permissions and Copyright

Permissions and copyright, and other IPR issues, have been causing difficulties for users of images from the outset of the digital age. This is particularly so in education, where there is an understanding of, and willingness to meet the principles behind such issues. The education community, whilst believing that staff and students should have free and open access to the images they need, also understands the need to ensure that the legitimate rights of the creator or owner are fully protected. This sets up something of a conflict that is hard to resolve; for example, in the arts and medical world, the rights of many images are necessarily restrictive, yet these same images can be of fundamental importance for the teaching of those disciplines. This issue is exacerbated when considerations are given to preservation and the longevity of permissions needed to ensure future access, particularly given that definitions of 'educational usage' will almost certainly change over time, and completely new needs will emerge that may be outside the permissions associated with the image. Nonetheless, issues for users regarding copyright of images preserved for long periods of time are, inevitably, similar to the issues of using digital images in general. The key to copyright or IPR issues regarding long term access can be summed up in two ways: all images within a preservation system must have complete permission clearance, for any current or possible future use, in perpetuity; or, there must be a built in facility or process that enables permissions to be revisited in response to user needs.

4.9.v Image formats

Formats do appear, on the surface to be of importance to users. However, when we dig a little deeper, this importance wanes. Most people seem to accept that new formats will come along and, on the whole, this is not seen as a problem; there is an implicit assumption, based on most people's experiences of digital technologies, that the new formats will be an improvement on the current, and that they will most likely be backward compatible – at least for as long as is necessary for content to be migrated forward in a suitable state. This is refreshing, and confirms the conclusions drawn elsewhere in this paper that is the component parts of the object that constitute its value, not the limited standard created by common usage of a particular format (e.g. TIFF). However, as with the technical issues of migration, whether the right components travel to the next format is dependant on the process used, the new format itself, descriptions added later and, of course, how much human intervention is applied.

[23] Nancy Brodie, 'Authenticity, Preservation and Access in Digital Collections', (Treasury Board of Canada Secretariat), http://www.rlg.org/en/page.php?Page_ID=243.

Furthermore, librarians, archivists and other conservators of analogue resources are, perhaps rightly, concerned that in the digital era format and media are increasingly hard to separate[24], as demonstrated in the 'What is an Image?' section of this paper. Consequently, the image format which might once have been described according to its media type (e.g. photographic negative), is now simply a binary code viewable only on a machine capable of understanding the particular variation of ones and zeros. As such the future of digital media does have an intrinsic difference to its analogue forebears: images preserved on paper, glass plate or even the humble thirty-five millimetre slide already have a proven track record and, in fact, still exist, where they have been properly preserved. Whereas, in the digital arena, much material has already been lost to the countless piles of 3", 3.5", 5", 8" disks and, as presentation and storage formats continue to evolve, it is a fair assumption that many digital images will be 'forgotten'.

As far as most user needs are concerned, it would seem that factors relating to format will ultimately be governed by the non-technical aspects of image quality, provenance and permissions and copyright, as identified above.

4.10 Taking user needs into account - recommendations

In summary, consideration for user concerns is dominated by worries, some founded and some speculative, relating to, first and foremost, access and the metadata required to facilitate access, followed closely by the difficult issue of copyright/IPR and the more transient nature of quality and changes in technology. But, in terms of what preservation systems must consider during their development and maintenance, it is the connection between *technological facility* and *human factors* that is imperative. Just as the different image formats develop, change or migrate into emerging formats, so a system must adapt and grow in response to issues relating to change of use, copyright, wider accessibility, and even metadata, within the communities that the system is designed to serve.

To respond to changes in an internal environment is, in some ways, an alien concept for a digital system; one of the overriding attributes of new ICT is its ability to automate change, to look after itself and reduce the need for human intervention. However, if preservation systems are to meet the needs of future users, perhaps some new ways of connecting system with environment need to be explored. This could work from *machine to human* with, for example, copyright information. The details of a licence could be stored, in explicit detail, within the metadata structure of a collection of images. This would allow the system to be more proactive with the possible uses, restrictions and conditions of a license whilst enabling it to automatically alert an appropriate person when human intervention is required; for instance, when a date-specific action is required. But, perhaps more importantly, preservation systems need to be considered in terms of how *human to machine* interaction can be improved; a system for preserving images is not a 'lock-it-up-and-bolt-it-down' facility for the protection of digital objects, it is a living process for guaranteeing enduring access to fit-for-purpose images for its users.

24 M.S. Lynn, 'The Impact of Digital Technologies in The Relationship Between Digital and Other Media Conversion Processes: A Structured Glossary of Technical Terms', http://palimpsest.stanford.edu/byauth/lynn/glossary/impact.html.

5. Properties of Digital Images

5.1 Introduction

File formats continue to evolve, becoming more complex as revised software versions add new features and functionality. It is not uncommon for software enhancements to render files generated by earlier versions unreadable. The threat to aging digital information has surpassed the danger of unstable media or obsolete hardware, the most pressing problems confronting managers of digital repositories are data format and software obsolescence (Lawrence *et al*., 2000, p. 1).

File (or data) formats define the rules used by application software to convert bits (the fundamental unit of digital data) into meaningful information that can be viewed and manipulated by a user. Most application software developers produce file format documentation for the formats they design and develop. Not all of them make this documentation available and even if they do, it is not always accurate (see Lawrence *et al.*, 2000, pp. 13-15 for examples of attempts to retrieve the Lotus 1-2-3 and TIFF file formats from their developers.).

Based on the availability and stability of the format specification, file formats can be classified as proprietary, open or standard formats. ***Proprietary file formats*** are not public and are developed and maintained by software producers. Larger software producers may sometimes publish their format specifications (PAS – Publicly Available Specification) or several firms may join together in a consortium to define interface standards so that they can develop mutually compatible products. These are called ***open or public file formats***. Some file formats are developed to become international standards (***standard file formats***) which are then public and fixed or stable until the next release of the standard. It is not unusual that software companies produce their own modified, proprietary, versions of standard file formats – these will be based on standards, but will have extensions that are proprietary and generally not public (e.g., Microsoft's version of XML). Many proprietary formats are, nevertheless, widely used and provide extensive compatibility with application software – these formats are often classified as *de facto* standards (cf. DLM 1997, pp. 50-52).

Successful and cheap long-term preservation of a digital file depends on the openness, level of standardisation and compatibility with other software products of the file format. Without a format specification the vital rendering tools that enable the use of digital files over longer time cannot be developed. Reverse engineering of software or the digital objects themselves can provide some answers, although legal constraints may well prevent this kind of action. Even where reverse engineering is possible, without any file format documentation, the process is likely to be too laborious and expensive (Leeds, 2003, p. 4).

The preservation risks associated with file formats are mostly related to loss of data and cost. Both migration and emulation — the two best digital preservation strategies currently in use — rely on file format specification being known and accurate. If it is not, the preservation strategies risk introducing distortion, loss of quality or data, or not being able to render the file usable at all. The risk management of file formats for preservation has to account for all these considerations.

This report considers both the familiar raster image and the perhaps less well known vector image. The former include the products of digital photography and scanning with file formats such as TIFF and JPEG. The latter may be less well known but in terms of digital images is probably more used as the almost ubiquitous PDF file, a subset of PostScript, is a vector-based format. Both can be said to be geometric or spatial but similarities end there. Raster images are grid-based with information being held about each point or pixel within the grid whereas vector images have information about any number X, Y, Z spatially defined coordinates. This chapter identifies the defining characteristics of both vector and raster file types, analyses and assess the principle file formats used in their creation, and provides information as to their suitability for long term preservation.

5.2 Vector Images: Overview

A vector image is created 'on-the-fly' either by plotting objects defined in a database or some other form of structured data, or by executing a sequence of drawing instructions. The image is rendered by software to a raster device such as a computer monitor or a printer. Vector graphics data are used in many drawing and illustration packages, CAD programs, most 3D modelling and animation programs, and coordinate based mapping applications. They are also often used in page layout and plotting situations. The definitions of scalable fonts available for printers and windowed computer environments also consist of vector data although these would not normally be considered as images.

Vector data may be referenced according to fixed, arbitrary or relative coordinate systems. Fixed coordinate systems are generally used by illustration and drawing packages with the coordinate system usually mapping to a page. CAD packages use arbitrary coordinate systems that suit the data. The basic unit of the coordinate system can be as small as an Angstrom unit or as large as a light year, and objects can exist anywhere within the limits of the numeric precision used to store the coordinates values (essentially from minus to plus infinity). Relative coordinate systems are used where only the relationships between the elements of the image are significant, for example in molecular diagrams, and also by some metafile formats which just need to know how far and in what direction to draw a line or curve from the current drawing position. Coordinate systems can theoretically have any number of dimensions but normally two or three are used.

The amount of descriptive information held in a vector data file varies between applications. Usually CAD files fully describe geometrical objects (e.g. lines, polygons, ellipses) and their attributes (e.g. colour, line type) along with their position and extrusion in two- or three-dimensional space. In certain applications the type of the objects is known and only their absolute or relative positions need to be defined as, for example, with GIS data where the objects are polygons (points and lines can be considered as simple polygons) and only lists of vertices need to be stored. Similarly with molecular or crystallographic data only the details of the chemical bonds needs to be stored as the rendering program can supply the atoms and representations of the bonds between them.

A consequence of storing the image information as vector data is that the coordinate system of the image must be mathematically transformed to map it to that of the target device. This transformation generally requires one or more of translation, scaling and rotation. As a result it is a simple matter to apply any level of zoom, select any part of the image to display and apply arbitrary rotations without loss of image quality by manipulating parameters in the transform. Vector graphics formats often store bitmap information. The bitmap graphic is generally stored as a simple rectangle having a complex fill (i.e. a raster image is embedded within a vector graphic).

5.3 Vector Images: Significant Properties

Significant properties of vector images differ depending on the general application and on the specific purpose of the data. A large array of different technologies and applications produce vector graphics with a consequence that the following guidelines focus on the generic issues. In some cases more specific information is available in the AHDS Guides to Good Practice series including guides to CAD, GIS, Geophysics, Remote Sensing and Virtual Reality[25].

5.3.i File formats

The file format determines which software is able to render the content to an output device. There is a degree of interoperability between software and file formats for given application types through the agreement and use of standards. These may range from agreement amongst software vendors to support open source export or exchange formats to the increasing support for XML within applications. In being ASCII based the latter provides an ideal basis for long-term preservation. Many proprietary vector graphics software packages use binary formats natively which can be problematic even when the format is documented as an open standard in that archiving in the long term involves version migration as the format evolves. Fortunately, many software packages (but by no means all) provide export facilities as ASCII text which as already noted is the ultimate preservation format.

5.3.ii Scale

In CAD and GIS files the creator decides on the basic unit of measurement. For geographic data this will usually be a metre, for architectural data either a millimetre or, in the U.S., a foot. It is important that the value of the basic unit is documented. Crystallographic and molecular data files fix the basic unit as being an angstrom. Scale is normally irrelevant in illustration files as the image is intended to fit on a page.

5.3.iii Precision and accuracy

The precision of the data in vector images is dependant on the discipline from which the data is drawn. Geospatial CAD and GIS data, when collected by survey, are likely to have been collected at a precision greater than necessary. Modern survey equipment usually works to a precision of a few millimetres whereas geographic data seldom requires better then a metre precision. Architectural drawings are generally drawn to a precision of a millimetre. Engineering data often requires sub-millimetre precision.

The accuracy of the data is, in many cases, important, and may even be critical in engineering drawings. In this case there should be a statement specifying the

[25] http://www.ahds.ac.uk/creating/guides/

reliability of the data, particularly if the data has been modified from its original format (e.g. through coordinate transformation).

5.3.iv Coordinate systems

Geographical coordinates can be expressed as angles of longitude and latitude or within a national or international planar Cartesian coordinate system. The former locates features on the globe whilst the latter within a map projection. Because maps are a two-dimensional representation in part of something spherical, distortion is generally a problem. Consequently, to reduce distortion coverage of large areas tends to be achieved through a large number of map projections covering smaller areas. Locations can also be covered by multiple projections.
GIS files and CAD files depicting geography are frequently geo-referenced to a particular map projection. In such cases it is vital that the coordinate system or map projection the data conforms to is fully documented. Clearly this is not the case for arbitrary or relatively referenced vector data.

5.3.v Geometry

The most significant content of CAD files is the geometry; the disposition of lines, points, text, and other simple or complex drawing objects in the two- or three-dimensional space of the model. For topographic data in GIS files it is the coordinate location of spatial units and what they are (i.e. points, lines or polygons) that is important. Other packages producing vector data rely on regularly collected point data in the form of a grid and thus have a fixed geometry.

5.3.vi Objects and their relationships

Crystallographic and molecular data define the atoms making up a compound and also the relationships between each atom as a description of the atomic bonds.

5.3.vii Conventions

Extra meaning can be given to objects in CAD files through the use of layers, colours, line weights and line types and also their incorporation into groups or blocks. These can be used to classify objects into groups or different types. GIS applications may use colour coding to represent different categories of information. Both CAD and GIS support layers. For example, one GIS layer may contain data pertaining to road systems whilst another might represent river systems. Layers can be viewed independently or superimposed. One potential problem area relates to invisible material included within layers (possibly switched off or frozen, etc). These can be purged unless there are clear reasons for retention. As such, conventions convey meaning and it is important that they are fully documented.

5.3.viii Associated data

Most CAD packages allow external data to be linked to drawings. These may be supplied with the software package, e.g. line styles and symbol libraries, or may be custom material. The latter can be user-defined line styles, symbols, bitmap images, or complete drawings, generally incorporated as blocks. Similarly some molecular formats have associated files. The absence of any such external data may not be obvious to a re-user and may substantially degrade the value of the image.

Generally speaking vector image formats concerned with geographical survey (geophysical, lidar, marine, etc) and a number of GIS formats can have associated non-geometrical data (attribute data). It is also possible to link non-geometrical data to many CAD file formats giving such files basic GIS functionality. CAD files may have databases holding data describing physical characteristics of objects or materials in the model or drawing or, particularly for architectural drawings, part identifications and costs for repeated objects allowing ordering information or bills of materials to be calculated. Illustrations are frequently associated with some text in an associated text document. In the absence of this text the image may be meaningless. Clearly the existence of associated data needs to be recorded and the data maintained as part of any long-term preservation strategy

5.3.ix Methodology

The methodology used in collecting survey and molecular data has substantial implications for accuracy and interpretation. Thus a GIS survey carried out using a total station EDM with an accuracy measured in millimetres is more trustworthy than one carried out using 30m tapes and triangulation, X-ray diffraction studies on a molecule may lead to a different interpretation than does nuclear magnetic resonance.

5.4 Vector Images: File formats

File format and extension	Description	Assessment of preservation suitability and risk
PostScript and its variants		
PostScript .ps PDF .pdf EPS .eps Illustrator .ai	PostScript is a page description language developed by Adobe Systems which specifies everything to be printed, including text, in terms of straight lines and cubic Bézier curves. This allows scaling and rotation plus other transformations to be performed on the content. Rasterizaton is performed on the fly either by dedicated software on the printer or by a program on the host computer that may either display the PostScript content on screen or send the rasterized copy to a printer. Encapsulated PostScript is a version of PostScript that describes part of a document and	A PostScript document is a 7-bit ASCII file but it is also a computer program. Like all computer programs PostScript is vulnerable to crashing. In being an interpreted program there is no guarantee that different interpreters will produce identical results[26]. Clearly the above also applies to derivatives The AHDS Guide to Good Practice Creating and Documenting Electronic Texts notes "…it is clear how programs like PostScript and PDF fall into the category of a

[26] Witten, I.H. and Bainbridge, D. (2002) How to Build a Digital Library. Morgan Kaufmann

	defines the bounding box that contains the part. EPS as a format is designed to be contained within another PostScript document. Many publishers favour EPS for the submission of line-art. Portable Document Format is a subset of PostScript in which a number of the programming elements such as 'if' and 'loop' are absent essentially just leaving the description of the page image. This requires much less processing and hence a much simpler interpreter to render the image to the screen or printer. Recently there have been further modifications aimed at making PDF more accessible to users with disabilities, for example tags, text equivalents (c.f. the HTML 'alt' tag for images), and audio descriptions can be included in the PDF file which can then be used by PDF readers to allow resizing and reflow of the text, or to play audio equivalents. Adobe Illustrator .ai files are PostScript files and by simply changing the file extension to .ps can be read directly by most software that can render PostScript.	proprietary processing language concerned with presentational rather than descriptive mark up. This does not imply that these languages should be avoided. On the contrary, if the only concern is how the document appears both on the screen and through the printer, then software of this nature is appropriate. However, if the document needs to cross platforms or the project objectives require control over the encoding or document preservation, then these proprietary programs are not dependable." which suggests unsuitability as preservation formats. The emergence of the PDF-A (for archiving) open standard suggests the above is not an entirely universal view. PDF-A is a constrained version of PDF 1.4[27] PostScript was widely used within the Publishing community but is being supplanted by the almost ubiquitous PDF
CAD		
WebCGM .cgm	Computer Graphics Metafile has been an International Standard Format (ISO 8632) since 1987. It is a 2D format that can contain vector and raster information as well as text. WebCGM is subset (or profile) of this standard to extending CGM as a suitable format for use in web documents through the incorporation of hot-spots and hyperlinks. WebCGM is a collaborative development between the World Wide Web Consortium and CGM Open, a section of OASIS (Organization for the Advancement of Structured Information Standards).	W3C standard which is seeing some take up amongst technical illustrators.
Scalable Vector Graphics .svg	SVG is an XML mark-up format that describes two dimensional	XML based W3C standard which is suitable for long

[27] http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml

	static and animated vector graphics. It was created by the World Wide Web Consortium and is an open standard. Work is underway to transforms SVG into a tactile representation to make it accessible to the visually impaired[28]	term preservation of 2D vector graphics.
Initial Graphics Exchange Specification (IGES) .igs	IGES is a neutral data format designed for digitally exchanging information between CAD/CAM (Computer Aided Manufacturing) systems. It is a public domain ANSI standard[29].	IGES is well proven but is voluminous and complex with the consequence of limited support from CAD vendors. Third party proprietary translators are available[30]. IGES seems to have had some uptake within the mechanical engineering community where the specification is targeted.
Standard for the Exchange of Product model data (STEP)	STEP is a neutral data format designed for digitally exchanging information between CAD systems. It is a multifaceted standard designed to cover most engineering activities associated with the manufacture of products[31]. STEP uses the EXPRESS modelling language to define models. The STEP format holds more than just geometry and can contain information to describe the complete history of a product from its inception to the final production[32]. STEP and EXPRESS are both ISO standards (ISO 10303 and ISO 10303-11).	As noted STEP is an ISO standard for the exchange of CAD files. As such it has the potential to act as a preservation format It appears more generic in not being so targeted as IGES (see above) but like the latter has seen limited support from CAD vendors. It is, for example, supported in AutoDesk Inventor and Mechanical Desktop but not directly in AutoCAD Drawing packages. Again, Third party proprietary translators are available.
AutoCAD Drawing Format .dwg	DWG is the AutoCAD native file format. Autodesk does not publish specifications of the format. The DWG format has changed substantially over the years reflecting improvements made to AutoCAD and also changes to the database holding the geometric and non-geometric data. A number of other CAD program authors use DWG as their format	In being a proprietary and closed format DWG would appear unsuited as a preservation format. It is also a binary format which is usually seen as not suited for long term preservation

[28] http://www.svgopen.org/2004/papers/TactileAccessToSVG/
[29] http://www.nist.gov/iges/
[30] http://www.tenlinks.com/CAD/translation/software.htm
[31] T. Barry, and K.W. Reynard, 'Computerization and Networking of Materials Databases' (ASTM International, 1992).
[32] ISO 10303 – Wikipedia http://en.wikipedia.org/wiki/STEP_(ISO_10303)

	but because the DWG format can only be determined through reverse engineering the programs usually use a DWG version some steps behind the current version of AutoCAD.	
AutoCAD Drawing Exchange Format (DXF) (also known as Drawing Interchange Format) .dxf	DXF is a CAD data file format, developed by Autodesk as a solution for enabling data interoperability between AutoCAD and other programs. Autodesk publishes the specifications of the DXF format but some recently added AutoCAD objects are not fully documented thereby reducing the usefulness of this format[33]. DXF files can be created as either ASCII text or binary files. DXF is intended to be an exact representation of the AutoCAD drawing and hence is equally vulnerable to changes resulting from improvements made to AutoCAD.	DXF benefits from supporting export as either ASCII or binary with the former seen as a stable media for long term preservation; however, failure by AutoDesk to keep the DXF specification up to date is problematic for other CAD vendors trying to provide support. Has long been seen as a *de facto* standard for CAD[34] in light of suitable open standards. STEP and IGES (see above) may come to fill this role. In the meantime an option is to practice version migration using the DXF format.
AutoCAD Design Web Format (DWF) .dwf	DWF is a compressed binary file of 2D or 3D vector data exported from a DWG file. It is a read only format designed originally for web delivery of 2D models but has been upgraded to contain 3D models. A special, free, viewer is required to view the models in the files.	This binary format requires a special viewer and is not suitable as a preservation format.
Hewlett Packard Graphics Language (HP-GL) .hgl or .hpgl	HP-GL is a printer control language developed by Hewlett Packard for their plotters that eventually became a standard for most plotters. It consists of a series ASCII drawing commands. It is used by the Engineering Support & Technology Division of CERN to archive their CAD drawings[35]. It offers the advantages of being software independent, being unencumbered by the non-visual elements of CAD drawings and containing only those elements of a drawing that are on visible layers. It is restricted to two dimensions.	HP-GL in being ASCII based is potentially a suitable preservation format but has only limited support for much CAD functionality.

[33] AutoCAD DXF – Wikipedia http://en.wikipedia.org/wiki/ASCII_Drawing_Interchange_file_format
[34] R. Walker, (ed.), 'AGI Standards Committee GIS Dictionary', (Association for Geographic Information, 1993).
[35] http://est-div.web.cern.ch/est-div/CAD/CDDSIBELIUS/sibcsenl.pdf

Bentley Systems' Microstation[36] .dgn	DGN (Design) is the file format supported by Bentley Systems' Microstation and Intergraph's Interactive Graphics Design System (IGDS) CAD programs. It exists in two versions: the Intergraph Standard File Format (ISFF) specification, published in the late 1980s by Intergraph, sometimes referred to as V7 DGN, or Intergraph DGN; and, since 2000, Bentley Systems updated version of DGN which includes a superset of DGN's capabilities, but which has a different internal data structure. It is properly referred to as V8 DGN. Both versions of the DGN file format are open formats and are documented by their developers. DGN files are binary files with variable-length records for graphic elements, and non-graphic data.	In being binary formats V7 and V8 DGN are only suited to short term preservation unless version migration is practiced (i.e. as software/format develops DGN files are migrated to a recent version).
OpenDWG[37] .dwf	OpenDWG is an open file format for CAD files that attempts to be compatible with the AutoDesk DWG file format. This is accomplished through reverse engineering DWG files but is only partially successful.	These, like AutoCAD DWG, are binary files. The reverse engineering currently only covers up to AutoCAD 2002 and is thus several versions behind.
Universal 3D File Format[38] .u3d	U3D is an open binary format under development by the 3D Industry Forum in collaboration with Intel and ECMA International that is designed to support the reuse of 3D CAD data. The goal is to produce an extensible 3D data format that enables compression, web streaming, level of detail control. Being an open format it may be viewed on non-proprietary software and hardware platforms. Right Hemisphere has licenced its 3D viewing technology based on U3D to Adobe for inclusion in version 7 of Acrobat and Reader[39].	Open but binary which limits its usefulness as a long term preservation format.
3D Modelling / Animation		
Virtual Reality Modelling Language .wrml .wrl	VRML is an ASCII file format that describes three-dimensional interactive vector graphics. Has been largely superseded by (or more correctly incorporated into)	In being ASCII based VRML is suitable as a long term preservation format but now largely superceded.

36 http://www.bentley.com/en-US/Products/MicroStation/Features+and+Benefits/DWG_DGN.htm
37 http://www.opendwg.org/homepage.htm
38 http://www.ecma-international.org/publications/standards/Ecma-363.htm
39 http://www.3dif.org/modules.php?name=News&file=article&sid=1055

	X3D (see below)	
GeoVRML[40] .wrml .wrl	GeoVRML aims to enable geo-referenced data, such as maps and 3D terrain models, to be viewed over the web by a user with a standard VRML plugin for their web browser. It extends VRML97 giving support for non-Cartesian coordinate systems and viewing and interacting. Like VRML has been incorporated into the X3D specification (see below)	As VRML
X3D[41] .x3d	X3D is the successor to VRML and is governed by the Web3D Consortium. It is the ISO standard for real time 3D graphics. It is based on XML and is backwardly compatible with VRML and GeoVRML. It is available in three encodings; UTF-8 (which provides the VRML97 compatibility), XML and binary[42].	In being XML based which in turn is ASCII compliant it is suited as a format for long-term preservation.
3D Studio Max[43] .3ds .ase .asc .map .max	3D Studio is a 3-dimensional vector graphics and animation program, written by Autodesk Media & Entertainment. It is one of the most widely used 3D software packages.	Proprietary but appears to support both binary and ASCII export. Specification not released in full. Unsuited as a preservation format.
Adobe (formerly) Macromedia Flash[44] .swf	Flash is a multimedia format (often referred to as a 'movie') that can contain vector and raster graphics and streaming visual and audio. The flash .swf file format is open, although binary[45], allowing anyone to export Flash movies. The Macromedia licence specifically disallows using the specification to create swf players.	Proprietary but published binary format not suited for preservation
Illustration / drawing		
CorelDRAW	CorelDRAW is a vector graphics editor developed and marketed by Corel Corporation.	It is an undocumented proprietary format unsuited for preservation.
Cartography and Geographical Information		

[40] http://www.geovrml.org/
[41] http://www.web3d.org/
[42] V. Geriomenko, and C. Chen, C., 'Visualizing the Semantic Web', (2003).
[43] http://usa.autodesk.com/adsk/servlet/index?id=5659302&siteID=123112
[44] http://www.macromedia.com/software/flash/flashpro/
[45] http://en.wikipedia.org/wiki/SWF

Systems (GIS)		
Geography Markup Language[46] (GML) .gml	GML utilises XML to express geographical features. It can serve as a modelling language for geographic systems as well as an open interchange format for geographic data. It is an ISO standard (ISO 19136) and is built on a number of other ISO standards collectively known as the 19100 family[47] (Fadaie and Kresse, 2004). GML is defined by the Open Geospatial Consortium.	In being an XML based schema and an ISO standard GML is very suitable as a preservation format for Geographical data
National (or Neutral) Transfer Format (NTF)	NTF, administered by the British Standards Institution, was designed for the transfer of spatial information. It can contain two- or three-dimensional geometry and also has facilities to enable the inclusion of the data quality as well as attribute descriptions, feature classifications and code lists[48]. It is the standard transfer format for Ordnance Survey digital data	Integral support for NTF seems limited but translators exist to migrate to other spatial formats. A general feeling is that its use is waning in favour of GML[49]. Even recent OS products are using alternatives.
ESRI ArcView Shapefiles[50] .shp, .shx	Shapefiles is an openly published format. It stores non-topological geometry as part of a set of data files making up a spatial dataset. It must be accompanied by in index file (.shx) and a dBASE file that holds the attributes of the shapes in the shp file.	This is an open but binary format which might be seen to limit its usefulness as a long term preservation format; however, like AutoCAD's DXF(see above) ESRI's SHP and E00 formats are generally accepted as *de facto* standards in light of alternative standards. Recent support within ESRI products for GML and WMS (XML based Web Map Service)[51] through add on modules is looking set to change this.
ArcInfo Coverage Export format .e00	The ESRI E00 interchange data format combines spatial and descriptive information for vectors and rasters in a single ASCII file. It is mainly used to exchange files between different versions of ARC/INFO, but can also be read by many other GIS programs. It is a common format for GIS data found on the Internet.	In being open and an ASCII based format E00 is more suited to acting as a preservation format. See ArcView Shapefiles above for notes about support for GML.

[46] http://opengis.net/gml/
[47] W. Kresse, & K. Fadaie, K., 'ISO Standards for Geographic Information' (2004).
[48] S. Dowers, 'Data Models, Representations and Interchange Standards In Parallel Processing Algorithms for GIS' ed. Healey, R.G., Dowers, S., Minetar, M.J. and Gittings, B. Taylor & Francis
[49] http://www.geoconnexion.com/magazine/article.asp?ID=816
[50] http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf
[51] http://www.esri.com/software/standards/product-support.html

MapInfo Interchange Format .mif .mid	MIF/MID is MapInfo's standard format[52], but most other GIS programs can also read it. The format holds three types of information: geometry (geography), attributes and display. The MIF file contains the geometric data whilst the MID file header and attribute data as delimited text.	Like the ArcInfo export format this format is ASCII based and open and thus a possible preservation format. Similarly MapInfo products provide support for GML and WMS Which are likely to be even more suited[53].
Chemistry/Crystallography	There is a wide range of digital formats for describing chemical and crystallographic data, so much so that translators between these formats are available[54]. A few formats are available over the internet and browser plug-ins or standalone programs are available to render these structures in 3D.	
Crystallographic Information File (CIF) 55 .str (CIF files equate to a suite of files[56] contained within a Self-Defining Text Archive and Retrieval (STAR) file	The Crystallographic Information File is an ASCII text format designed for the electronic transmission of crystallographic data[57]. It has been adopted by the International Union of Crystallography as the recommended medium for this purpose.	In being ASCII based and having been adopted by the IUC CIF is a reasonable candidate as a preservation format but may have been overtaken by events and the emergence of CML (see below).
Elsevier MDL Molfile Molecule File .mdl	MDL Molfiles hold information about the atoms, bonds, connectivity and coordinates of a molecule. It was primarily designed as a format to submit molecular descriptions for publication. Most chemoinformatics applications can read this format. An XML version has been developed.	Not designed as a preservation format but as a submission format for print publishing. Potentially in its XML form it could act as a preservation format.
Brookhaven Protein Databank File58 .pdb	The protein databank file format coordinates of atoms in a molecular structure and information about the chemical bonds, as well as bibliographic data. It is a widely used format for describing macromolecules. PDB files are plain text (ASCII) files. An XML version has also been produced[59].	Again ASCII based but specific and thus limited.

52 http://extranet.mapinfo.com/common/library/interchange_file.pdf
53 http://mapinfo.com/location/integration?txtTopNav=NOT_SELECTED&txtLeftNav=NOT_SELECTED&txtExtNav=NOT_SELECTED&txtDetailType=FREEFORM_TEXTAREA&txtDetailID=434
54 http://openbabel.sourceforge.net/babel.shtml
55 http://www.iucr.org/iucr-top/cif/spec/version1.1/index.html
56 http://tang.bmrb.wisc.edu:8080/WebModule/wattos/MRGridServlet/html/readme.html#List_of_file_extensions
57 W. Clegg, 'Crystal Structure Analysis', (Oxford University Press, 2002).
58 http://www.rcsb.org/pdb/docs/format/pdbguide2.2/guide2.2_frame.html
59 http://pdbml.rcsb.org/

Chemical Markup Language (CML) 60	CML is an XML format to describe molecules and substances and their physical properties and also chemical reactions and analytical information, such as spectra. It was designed as an exchange and archive format, but as with any form of XML, the rigid structure of the data enables software to extract and render full 3D models from the contents.	In being XML based it will probably become a stable preservation format for chemical relationships.
Other Vector Graphic formats		
	There are large numbers of small usage, largely proprietary software packages supporting vector graphics and used, for example, in laser scanning and terrestrial and celestial survey. Many of these technologies pertain to so-called 'Big Data' in that datasets tend to be exceptionally large. The Archaeology Data Service (who host AHDS Archaeology) is currently undertaking a project for English Heritage looking at management and preservation strategies for such data, including a format review[61].	Disciplines using these packages include the Earth Sciences, Oceanography, Maritime studies, Astronomy, Archaeology and Geography. Many of these packages use binary as a native format. Some, but by no means all, also support ASCII exports. The latter provide the best option in terms of preservation.

5.5 Vector Images: Risk assessment and Recommendations

Many vector graphics software packages generate binary files which can be problematic for longer term preservation. Fortunately many, but not all, allow for export as structured ASCII text which is generally accepted as being the most stable of formats. An ASCII export is not enough in itself as its meaning has to be openly documented. Formats that either export or natively use ASCII including XML are inherently stable. In general if long-term preservation is an issue software packages that only support binary should be avoided. The following commonly used technologies and formats therein are recognised as having specific problems.

5.5.i PostScript and its variants

PDF appears to be subsuming other PostScript formats. Adobe have released the documentation for this popular format as an open standard[62] but the preservation of PDF files has always been an issue. A frequently used solution is to capture each

[60] http://wwmm.ch.cam.ac.uk/moin/ChemicalMarkupLanguage
[61] http://ads.ahds.ac.uk/project/bigdata/
[62] http://partners.adobe.com/public/developer/pdf/index_reference.html

page of a PDF in an alternative format such as a raster format such as tiff. Recently, an international standard, ISO 19005-1, has been agreed for the long term preservation of PDF files. PDF-A (for archive) is based on version 1.4 of the PDF format but is constrained in that, for example, executables including use of JavaScript are not allowed, fonts must be embedded and encryption is prohibited[63]. That a PDF file will still contain executable instructions (PostScript) there are still questions about its suitability as a preservation format.

5.5.ii Computer Aided Design (CAD)

The text-based version of DXF format used for exchanging files between CAD (Computer Aided Design) packages is maintained by the major CAD software vendor, AutoDesk, and reflects the latest version of their package AutoCAD. AutoDesk have not released documentation about changes in recent versions of AutoCAD which means other vendors cannot support them. In short it is no longer an open standard. Consequently, the only realistic way of preserving AutoCAD drawings is through version migration. Purchasing and migrating through successive versions can be a very expensive procedure in terms of time as well as financial outlay.. As noted above STEP is an ISO standard for the exchange of CAD files but there has been only limited physical implementation thus far.

5.5.iii Geographic Information Systems (GIS)

The major GIS software vendors, ESRI and MapInfo provide import and export routines for the interchange of data. The lack of an open standard has led to ESRI formats being accepted as /de facto/ standards amongst users. Recently ESRI have introduced add-on modules for their ArcGIS and ArcIMS packages which provide support for XML based open standards; Web Map Server (WMS) and Geography Markup Language (GML)[64] and hence compliance to the Open Geospatial Consortium (OGC) standard[65]. A JISC funded Interoperability Demonstrator Project is currently investigating the use of these standards.

5.5.iv Chemical and Molecular formats

The emergence of CML should cater for such formats.

---

[63] http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml
[64] http://opengis.net/gml/
[65] http://www.opengeospatial.org/

5.6 Raster Images: Overview

The raster image is the most common category of image created and used within digitisation projects, and delivered over the Internet.[66] All scanners and digital cameras produce raster images and most output devices (print and screen) also use them.

Raster images take the form of a grid or matrix, whose pattern becomes easily visible as the image is magnified. See fig 1:

Figure 1: Diagram showing a raster image zoomed to display its pixel

Lucie Rie, Ceramic Buttons, c.1940s, © Mrs. Yvonne Mayer/Crafts Study Centre 2004

Each square or picture element (pixel) within the matrix occupies a unique position and can be edited separately. Each pixel stores colour information and the pixels taken in their entirety make up the image. Internally, a raster image's coding typically includes a header describing the structure of the file followed by a series of values, each describing the colour of the individual pixels.

Since raster images record information for each pixel, their file size can be relatively large. For an uncompressed raster image, the file size will be directly related to its pixel dimensions (spatial resolution) and the extent of the colour information recorded for each pixel (its colour resolution or 'bit-depth').

Although most raster file formats are similar in structure, they can be distinguished by the amount of information they record per pixel (i.e. their bit-depth), the methods used to record their code more efficiently (their compression), and the additional functionality they offer (e.g. transparency layers, colour management or metadata support). They can also be divided into open formats and proprietary formats.[67]

The main defining features of a raster image are:

The file format of the image relates to the type of computer code that is used to structure the raster image. There are many choices for the type of format to choose for raster images. Different file formats offer differing levels of compatibility and compression. For archival purposes non-proprietary formats without compression are best.

[66] This is an edited version of the article: *File formats and compression* found at the TASI website http://www.tasi.ac.uk/advice/creating/fformat.html
[67] see TASI http://www.tasi.ac.uk

5.6.i Resolution

The resolution of an image concerns the number of pixels held within the digital file, and is measured in pixels per inch (ppi). The more pixels stored per inch, the greater the density of the colour information, and therefore the greater the detail evident in the image. The appropriateness of the resolution chosen depends on the intended purpose of the digital image. However, note that resolution or ppi is only an indicator of image size, and therefore 'quality', when we know the dimensions of the original analogue object. For example, scanning an A4 document (9 × 12 inches) at 300 ppi will produce a digital image that is 2700 pixels x 3600 pixels (the dimensions of the original multiplied by the ppi). Scanning a postage stamp that is 1 inch x 1 inch in size, will produce a digital image that is 300 pixels x 300 pixels. Both these images are scanned at the same 300ppi resolution, but produce vastly different sized digital images. A more accurate way to refer to the size of a raster image therefore is to use its pixel dimensions.

5.6.ii Bit Depth

Bit-depth refers to the amount of colour information held with each individual pixel. In a greyscale image there are usually 8 'bits' of information in each pixel, in a colour image 24 bits are standard, although some software will scan at 48 bit. The number of bits of information held in a raster image also impacts on file size.

5.6.iii Colour Space

The colour space of an image refers to the method of working with colour based on colour models. The most common colour models are bitonal, grayscale, Indexed colour, RGB and CMYK. The bitonal mode uses only two colour values, black and white, to display images. The grayscale mode offers 256 shades of grey that range from white to black. Indexed colour is the limited palette of 216 colours that web designers are able to use which display on both Macintoshes and PCs. It is more usual to work in RGB and then only convert to Index colour mode if you think it will really effect your users e.g. if a logo has to display as a certain shade of red. Computer monitors and TV screens use the RGB model to display a mixture of red, green and blue colour values. The CMYK model refers to the printing colours of cyan, magenta, yellow and black.

The colour model effects the file size of the image, since the more bits and bytes, the larger the file size. Images with only black or white pixels (bitonal – 1 bit per colour channel) will therefore have the smallest file sizes, grayscale images quite small file sizes (8 bits per colour channel) whereas RGB images with 24-bits (8 bits per colour channel) or more will have much bigger file sizes.

Another issue to consider is that it can be difficult to convert from RGB to CMYK. This is normally done when you need a commercial company to print an image for you. When you convert to CMYK you need to be aware of the 'gamut'. The term gamut refers to the range of colours that the combination of CMYK inks print. Some colours may be 'out of gamut' and therefore can't be printed accurately. This can be resolved with certain programmes, such as PhotoShop, that provide an 'out of gamut' warning, and give you options to replace a colour with one that is in the gamut. It is best to work in RGB and then keep a copy as RGB and convert another copy to CMYK if needed for printing purposes.

5.7 Raster Images: File Formats

This section will provide most information on and analysis of the four main contenders as stable long term raster image preservation file formats – TIFF, PNG, JPG2000 and DNG, giving information on origins, characteristics, suitability and any potential shortfalls both now and into the future. Following this discussion, other proprietary formats and those less suitable for preservation will be listed and a short description attached where appropriate.

File format and extensions	Description	Assessment of preservation suitability and risk
Tagged Image File Format (TIFF)	The TIFF file format is to date the most common and widely used preservation image format. Often it is referred to as a 'de facto' standard, in that the format was never explicitly intended as an image preservation standard, but has been taken up by the industry as such. TIFF is a proprietary format – developed commercially by Aldus and Microsoft, and now maintained by Adobe. The format is flexible enough to be enhanced and amended by software and image processing vendors as they see fit, and was always intended as an extensible format to allow for optional functionality and extension.	This extensibility presents some problems for image preservers, as there are now various types and flavours of TIFF available, and as a result there have been some famous mis-matches. For example, TIFF images made by SCITEX products - one of the largest makers of scanners in the world - were for a while unreadable by Photoshop - the main image editor. In addition, TIFF compression (LZW owned by UNISYS) is based on proprietary patented software and is unsuitable for long term preservation. There is only one flavour of TIFF suitable for long term preservation: un-compressed Baseline Revision 6. It should not use any compression and should not use any of the additional functionality available in some other revisions.
Jpeg 2000	JPEG2000 has been an international standard since December 2000 when JPEG 2000 Part 1 became ISO/IEC standard 15444. The following year an extended version .jpx or 'JPEG 2000 Part 2' reached ISO Standard status. The format offers lossless compression or high quality wavelet[68] based lossy compression, and has the possibility of massive bit depth: 214 channels, holding up to 38 bits of information each, far more than any image editor can handle at present. JPEG2000, also provides several progressive display options, such as	JPEG 2000 could be the format of the future, effectively allowing the same file format to be used for 'master archive' and 'surrogate delivery' files. JPEG 2000's compression is more efficient than other common compressions. It will deliver lossy images 3-5 times smaller than comparable JPEG images. Lossless JPEG 2000 images are necessarily larger, but are still generally half the size

68 Description of wavelet compression http://www.tasi.ac.uk/advice/creating/wavelet.html

	a rough, full size version that builds up in detail (i.e. display by fidelity), or a small image that grows until it reaches full size (display by resolution), in addition to the traditional pixel by pixel display of normal raster images.	of the original uncompressed raster image. This is better than the lossless LZW compression (used in GIF and optionally in TIFF) or Deflate (used in PNG). However, the format again has the baseline v's extended format differences problems associated with TIFF above, and has had limited support from the main browser manufacturers. Most browsers require a plug-in the view the JPEG2000 files. Similarly Adobe, have to date been sketchy in their support of JPEG2000, although latest versions (7 and above) do read/write the format. Moreover, at time of writing JPEG2000 still does not have the support of W3C. In short, JPEG2000 is likely to be a stable future archival preservation file format and offers excellent functionality in terms of viewing and compression, but at this time (November 2005) industry and browser support is still to a great extent lacking to enable a full endorsement.
Portable Network Graphics (PNG)	Developed as gif substitute and possible tiff substitute, PNG was developed in 1995 and is working towards ISO standardisation. It is a fully open source format, supported by W3C. It has a bit depth of up to 48-bit true colour (as opposed to 8-bit 256 colour GIF) offers patent free (unlike TIFF) lossless compression.	The main problem, in terms of preservation, with PNG is simply the lack of support and confidence that the industry and users have given it. The main culprit here may be Adobe. Photoshop has done very little to provide support for more than the most basic PNG functionality.[69]

		Also, Macromedia has done little to help. They have adopted the PNG as their internal default format in programmes such as Fireworks, but have chosen to add a range of additional functionality to the file type (objects, vectors, animations) which make the file unreadable in other programs, for example a fireworks PNG will crash IE. It is questionable if these formats are still PNG files at all, and they are they should not be considered baseline. Therefore PNG has the same problem as TIFF and to some extent JPEG2000, in that for preservation purposes only baseline versions of the file are suitable.
Digital Negative (DNG)	The DNG file format is Adobe's answer to consistent archival file formats. It is an attempt to standardise digital camera RAW formats, which currently are proprietary depending on camera manufacturer.	The fact that the format is being developed by Adobe, the leading image editor, is in its favour as this should ensure uptake and compatibility with all major browser and software manufacturers, and sufficient uptake from users. However the format is perhaps most suitable for images that have been created in a RAW format, such as those made by one-shot digital cameras. We will have to see if it becomes a standard archive format for all digital images.

		Adobe are also releasing a free software utility called the Adobe DNG Converter, which will convert many proprietary RAW format images into the new .DNG file format, compliant with the Digital Negative Specification.
Joint Photographic Expert Group (JPEG)	This is a popular file format for Web publishing and ideally suited to images where the picture is a photograph. However because JPEG uses "lossy" compression, which means that some of the data that makes up the image is lost in the compression process, it is not suitable for archival purposes.	JPEG uses "lossy" compression, which means that some of the data that makes up the image is lost in the compression process. It is not suitable for archival purposes.
Graphic Interchange File Format (GIF)	GIF is a popular file format for Web graphics and publishing, and is especially useful for images where the picture is line art, has large areas of solid colour or uses a limited colour palette.	GIF is not recommended as a preservation file format for digitised photographs, slides or other forms of complex images. However, GIF is a popular file format for Web graphics and publishing, and is especially useful for images where the picture is line art, has large areas of solid colour or uses a limited colour palette. GIF uses "lossless compression" (LZW), so there is no image degradation or blotchy colour.
MrSID (SID)	MrSid is a proprietary format intended for the storage and delivery of very large images, particularly geo-spatial images	At time of writing it is not suitable as an archival format for raster images but since it uses the same wavelet compression as JPEG 2000 it can potentially deliver the same functionality and may be re-assessed in the future.
Bitmap (BMP)	The standard graphics file format for MS Windows	As this is a proprietary standard graphics file format for Windows, Bitmap is not a preferred file format for long term preservation.
DjVu (DJVU)	DjVu is useful for compressing documents with a combination of text and images, rather than just individual images.	It is unsuitable as a preservation format as it is firstly proprietary and also because overall the compression is very lossy, which therefore involves degradation of the image.
PixelLive / VFZoom (PFZ / VFZ)	PixelLive is used to encode types of images typically stored as rasters, although the encoding used is vector-	Both formats use Genuine Fractals, which break the image into small shapes

	based. VFZoom is an earlier version of Celartem's PixelLive format.	(fractals), which are described mathematically and can be redrawn at a larger scale. Whilst they are effective as scaling and display formats they are not suitable for archival storage.
Photoshop Document (PSD)	The standard Adobe Photoshop file format. Used mainly for image files created in Photoshop i.e. images that are not a digital surrogate of an analogue object, but 'born digital' images.	When working in Adobe Photoshop it is possible to save the image in the proprietary PhotoShop Document format. This file format can also be converted into other file formats. Use of the this format is acceptable as the preservation copy of a born digital image. For any other raster image use, conversion to a more suitable preservation standard is recommended.

5.8 Raster Images: Risk assessment and recommendations

5.8.i TIFF

Baseline TIFF revision 6 is a recommended file format for raster image preservation. No compression or additional functionality, available in some other format revisions, should be used. Having been the 'defacto' standard for digital image preservation for the past ten years with widespread industry and software support, it is expected to remain an acceptable preservation format for the foreseeable future.

5.8.ii JPEG2000

JPEG2000 is, in its current form, an acceptable preservation standard. It has the advantage of ISO status, is non-proprietary and has some excellent functionality. However at time of writing the format falls short of full recommendation. This is mainly due to relatively slow take up among industry and leading software and web browser developers. Although difficult to predict, the likelihood, indeed the hope, is that over time this support will be forthcoming and JPEG2000 will achieve fully recommended preservation format status.

5.8.iii PNG

To some extent PNG is in the same position as JPEG2000 above, in that it is an acceptable preservation format but falls short, at time of writing, of fully recommended status. It is an open standard with W3C support and, as of 2003, ISO status. The main caveats to full recommendation are lack of browser support (Internet Explorer for Windows being the main culprit), and simultaneously a lack of full software support among some vendors, and some additions to the format made by others. As with TIFF above, only baseline versions should be used for image preservation. And as with JPEG2000 above, the hope is that the lack of support will be overcome over time, and PNG will achieve fully recommended status.

5.8.iv DNG

At time of writing DNG is not recommended or acceptable as an image preservation standard. It is a new addition to the range of possibilities, so is untested in its wider application, and may be limited in use to images created using one shot digital cameras. It is also proprietary, although the same caveat applies also to TIFF. Furthermore, as DNG is an Adobe developed format there is a strong possibility that wider software support will be forthcoming in the next few years, and the format's 'not recommended' status may need to be revisited.

6. Preservation Methods

6.1 Introduction

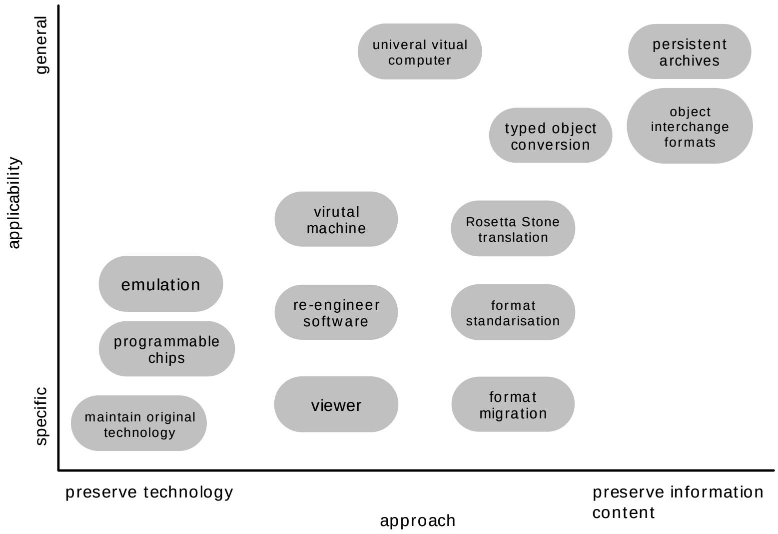
The great difficulty anyone trying to undertake preservation of digital content will encounter is the lack of practical advice, and of robust tools and mature techniques for digital preservation. A number of digital preservation strategies have been proposed, but there is no definitive approach to the problem of maintaining digital content across multiple generations of technology. Unfortunately, information on the likely costs, possible limitations and long-term sustainability of different strategies is still quite limited – partly for the very valid reason that no one has yet had the time to gain the experience needed to answer these questions.

Few organisations with digital preservation responsibilities appear to have yet fully developed their policies in this area (ERPANET, 2003). It is unwise to commit to a course of action before its consequences are clear, nevertheless it is equally important that organisations which make some claim to preserve digital resources should declare to their stakeholders what they can do to achieve this goal at the present time. Decisions about preservation methods might usefully take into account the following three-tiered understanding of digital preservation:

- Preservation of the bit stream (basic sequences of binary digits) that ultimately represent the information stored in any digital resource
- Preservation of the *information content* (words, images, sounds etc.) stored as bits and defined by a logical data model, embodied in a file or media format
- Preservation of the *experience* (speed, layout, display device, input device characteristics etc.) of interacting with the information content

Techniques for achieving the first of these objectives are well understood and include environmentally controlled storage, data replication, backup, and media refreshment. In the OAIS model, much of this activity falls into the archival storage function. The second and third objectives present a far greater challenge.

Binary data remains useful only for as long as it can be correctly *rendered* (displayed, played-back, interacted with) into meaningful content such as text, images and video clips. The process of rendering is performed by a complex mix of hardware and software, which is subject to rapid obsolescence. As a rule of thumb, it is reasonable to predict that current hardware and software will be able to correctly render a file for around ten years after its creation. By the end of this period, repositories need to have adopted a more active preservation strategy than simply preserving the bit stream of the file if they are to maintain access to information content held in the file. Either old data must be altered to operate in a new technical environment (migration, format standardisation) or the new environment must be modified so that it can render the old data (emulation, virtual computers). Within these two broad approaches there are many different techniques (figure 3).



Digital Preservation Strategies (based on Thibodeau, 2002)[70]

This chapter reviews some of these techniques and makes recommendations for approaches to take for both raster and vector images, and the preferred formats for preservation. It starts with a look at bitstream preservation and then spends some time looking at the preservation of content and the experience.

6.2 Overview and Assessment of Preservation Methods

6.2.i Bitstream preservation

This simply involves retaining the original data as a single sequence of binary digits (bits), i.e. the original data in an uninterpreted state. Bitstreams may be preserved in two ways: either as the original file in the data format as received, e.g. An MS Word document, a TIFF image, a plaintext file etc.; or in a normalised bitstream format, e.g. as a sequence of bits contained inside XML wrappers. Bitstream preservation is widely viewed as a form of 'insurance' in that it allows for the possibility of using future techniques for making content accessible. Moreover, it is an additional form of data backup.

It is hard to see that the first method offers any real advantages over the second. In either case metadata is needed to make sense of the file. Using the first method, the metadata would need to be kept separately but associated with the data content, perhaps, for example, in a separate METS file. The second method has the advantage that metadata about the file and its format can be included within the XML wrappers surrounding the content bitstream and is thus always kept with the associated content data[71]. The second method is to be preferred.

6.2.ii Preservation of Technology

[70] 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.
[71] See H. James et al, *Feasibility and Requirements Study on Preservation of E-Prints*, JISC 2003, 32-33

Technology preservation is an attempt to ensure the usability of digital resources over time by preserving and maintaining software (applications and operating systems) hardware, essentially creating an IT museum.[72] This may be a useful short-term solution but costs in terms of the maintenance and storage of equipment makes it prohibitive for any sustained strategy. As an example, AHDS Archaeology maintains a small 'computer museum' (http://ads.ahds.ac.uk/project/museum/) but 'exhibits' are intended to facilitate data recovery, not time-based access to source files, and have been used in earnest for recovery purposes. Data is moved into stable open formats and thus into a format migration strategy. At one stage the museum contained two working Amstrad computers but both of these have now failed and are not able to be repaired. A hardware preservation strategy would fail in the same way, without the expenditure of large sums of money to ensure that technologists with appropriate skills are always available. This approach cannot be regarded as viable.

6.2.iii Migration

Migration as a preservation term can be used to describe both file format migration and media migration. Media migration is more often known as 'refreshment' and is necessary to ensure that data is not lost through media degradation over time. The lifetime of media must be estimated and migration to new media undertaken before the threshold is reached.[73] Today there is a move toward storage on hard disk with multiple disk or tape backups. Integrity between versions is maintained through the use of fixity or checksum values as with the OAIS reference model. Refreshment is an essential activity in any preservation programme, but does not in itself ensure data preservation.

File format migration is used to ensure the accessibility of a digital object when the software it depends on becomes obsolete or unusable. It can involve conversion of digital objects from one file format to another (not necessarily the same) format, for example from Word 98 to Word 2000, from Word 2000 to Adobe's Portable Document Format (PDF), or from GIF to PNG. Some attributes of the digital object may be lost during the conversion process, so the experience may not be equivalent after migration. The level of data loss through migration depends on the number of preservation treatments applied to the record, the choice of process, the new data format, the level of human intervention and post-migration descriptive work.

Migration as an approach has a number of variations. The traditional technique is to migrate file formats to newer versions of the same format as the earlier versions approach complete obsolescence, eg Word 95 to Word 98 to Word 2000. This method requires migration at as late a stage in the life of the digital format as possible. A second migration technique involves migrating digital objects to limited range of standard formats at the time they are ingested by a digital repository. This involves fewer overall migrations but still requires the migration of the standard formats as existing versions of those formats become obsolete. A third technique depends on the migration of all file formats to a standardised file format which is chosen for its presumed longevity as a digital format, eg XML. This migration technique is often referred to as 'normalisation'.

---

[72] T. Hendly, "Comparison of Methods and Costs of Digital Preservation", *British Library Research and Innovation Report 106*, 1998, 16-17.
[73] T. Hendly, 1998, 12.

6.2.iv Emulation

Emulation is a technique often proposed as a solution to the hardware/software obsolescence problem. Put simply, emulation involves the development of software to replicate the behaviour of obsolete processes (such as hardware configurations or operating systems) on current hardware. Emulation thus aims to recreate part of the original process that interprets the data to produce a modern rendering of the original performance. Much emulation work is motivated by a belief that the original 'look and feel' of a digital resource must be maintained forever. 'Look and feel' includes the content of the record, but also tangible aspects of the presentation of the content, such as colour, layout, and functionality. However, it has been pointed out by a number of commentators that "traditionally, preserving things meant keeping them unchanged; however … if we hold on to digital information without modifications, accessing the information will become increasingly more difficult, if not impossible."[74]

The major problem with emulation is that it requires not only the retention of all relevant software applications (i.e. One for every type of file format being preserved), but also the coding of an emulator for every hardware and operating system in use. As well, the whole issue of the significance of the original 'look and feel' is a discussion studiously avoided by proponents of the emulation approach, but is assumed to be of prime importance. Many people state that recreation of the original experience is of prime importance for digital preservation yet there is little or no quantitative evidence to back this up, and the position is neither a given nor unarguable. It remains little discussed by proponents of emulation, however, but is an important underpinning/justification of this approach and needs to be debated widely.

6.2.v Migration on Request

Conceived by the CEDARS project, migration on request confusingly promotes emulation as a preservation strategy.[75] One of its chief developers says of CEDARS: "Original objects are maintained and preserved in addition to a *migration* tool which runs on a current computing platform. This would be employed by users to convert the original bytestream of the object they want to use into a current format".[76] A separate migration tool would be needed for each data format ingested by the repository and these would need to be maintained over time, ie. re-coded to cope with operating system and hardware changes. As well, this approach requires the development and maintenance of rendering tools so the migrated object can be viewed. Again, a separate rendering tool would be needed for each data format and these also would need to be updated as operating system and hardware changed. Software development times suggest that the rendering tools will need to be developed at the same time as the migrating tools so they are on hand to allow a migrated object to be usable at time of migration. The sustainability of such an approach is questionable.

Thus, the original object is maintained and the migration tool or emulator is migrated as computer platforms change. This is seen by the CEDARS team as more practical and cost efficient than the original model proposed by Rothenberg where, because of the expense, emulators were to be developed only when necessary. In the latter case a huge technological gap could build up, but the effort involved in migrating a

---

[74] Su-Shing Chen, "The Paradox of Preservation", *Computer*, March, 2001, pp. 2-6. [Professor Department of Computer Information Science & Engineering, University of Florida-Gainesville]
[75] See http://www.leeds.ac.uk/cedars/.
[76] P. Wheatley, "Migration - a CAMiLEON discussion paper", *Ariadne* vol. 29 2001, available at http://www.ariadne.ac.uk/issue29/camileon/.

tool from one version to the next as platform changes occur should be minimal. Clearly, however, this approach would require an ever increasing number of migration tools as software versions and formats increase. There is the further issue of who is going to create, migrate and maintain a repository of the tools. It seems unlikely that individual repositories could sustain such an approach without a large community of implementers of migration on request to call on for sharing of migration and rendering tools.

6.2.vi Universal Virtual Computer

A development related to emulation as a preservation strategy is the so-called 'universal virtual computer' (UVC). This concept was proposed in 2000 by Raymond Lorie of IBM, in research paper written for IBM, and later published more widely in an article in *RLG DigiNews*.[77] In brief, the UVC is a virtual representation of a simplified computer that will run on any existing hardware platform. Its appeal seemingly lies in the fact that problems of hardware and software obsolescence become irrelevant, and digital objects can be retained in their original format. It is said by its proponents to have the advantages of both the emulation approach and the format migration approach, with none of the disadvantages.

The only implementation of this concept is at the Netherlands Koninklijke Bibliotheek (KB) where a test implementation has been developed for preserving digital images (in fact PDF files, each of which is manifested as a sequence of JPEG files).[78] In this, the only real world application of the concept, it becomes clear that there is a large software development and maintenance load on implementers. In order to preserve JPEG images today the KB needs the UVC emulator, a format decoder (and will need one for each format being preserved), a logical data scheme, an equivalent of a document type definition or DTD, (again it will need one for each format), and a viewer that allows viewing of the decoded file. To access the file in the future the KB will also need to develop a UVC emulator for every hardware/software configuration on which the file will be accessed throughout its useable life.

The approach may have some worth, and the KB test implementation is regarded by them as demonstrating "the only method so far that guarantees long-term accessibility of PDF files". However, this is a sweeping statement which is not based on their own experience and which, in any case, can hardly be shown to be accurate except by the passage of time (and it is only a year since the original project finished). The KB implementation itself shows the problems of the UVC approach. It was set up to as a solution for preserving PDF files, but both KB and IBM found that developing a decoder for PDF was too difficult and complex a task to be completed during the life of the project. Instead, the PDF files were migrated to JPEG – a jpeg image for every individual page in each PDF file – and the JPEG files are preserved as the preservation version of the PDF files. The KB admits that "some of the original aspects of PDF publications are lost when using this method", which would seem to compromise the aim of the project "to preserve the original object".[79]

As well, comparisons of this approach with others in the KB document display a peculiar blind spot in advocates of the UVC: an inability to recognise that migration is

[77] R.A.. Lorie, "A Project on Preservation of Digital Data", *RLG DigiNews*, vol. 5 no. 3. At http://www.rlg.org/legacy/preserv/diginews/diginews5-3.html.
[78] See http://www.kb.nl/hrd/dd/dd_onderzoek/uvc_voor_images-en.html.
[79] Ibid.

not a single technique.[80] It seems likely that the software development burden, as well as issues around the complexity of developing decoders mitigates against the use of this approach in all except the most well-off or financially secure organisations,[81] or those with an extremely limited (and relatively simple) range of file formats being ingested.[82]

6.3 Summary of Image Basics

As described in Chapter Five there are two methods for representing the data that makes up an image: bitmap (raster); and vector. Bitmaps represent images as a series of picture elements (pixels) with associated information about colour values; vectors represent images as a set of coordinates and/or mathematical expressions that define the geometric shapes that make up the image. Each has its strengths and weaknesses. Raster images are excellent for representing continuous tone images (i.e. what we usually think of as photographs); vector images are much more useful for representing physical, scientific, and engineering images (e.g. from GIS and CAD systems), particularly because they are resolution independent, and resizing a vector image can be done without significantly affecting image quality.

Retention of source bitstreams is a necessary component of any preservation approach to safeguard against migration errors and choices of preservation formats which might prove to be incorrect over time. It is also important to bear in mind that preservation formats are not necessarily distribution / delivery formats (although dissemination / delivery versions of digital images may be in the same file format as the preservation copies).

Some other points need to be made:

- There is no **single** best way to preserve any digital resource;
- Decisions about preservation approaches depend on resources available, current and future use of the resources, and the cultural/historical/social/legal significance of the resources;
- Decisions made about the recommended preservation formats for individual resources can change over time.

6.4 Raster Images

In a sense raster images are a relatively uncomplicated digital type when it comes to considering preservation issues. Since they consist of a two-dimensional pixel grid

[80] The web page cited in note 3 assumes that migration is only a technique of continual same format migration, as do other papers by KB staff, and a paper by H.M Gladney in which cost comparisons are made. See H.M. Gladney, *Trustworthy 100-Year Digital Objects: Durable Encoding for When It's Too Late to Ask*, available at: http://eprints.erpanet.org/archive/00000007/01/TDO_Durable_final_submission.pdf

[81] The KB claim that “no periodic actions are required (unlike migration)” is demonstrably misleading, since the UVC emulator itself will need to be recoded every time the KB changes its hardware platform (or OS).

[82] A single example will suffice to demonstrate the burden such an approach could place on collecting institutions. The George Mason University (Virginia, USA) manages a 9/11 archive which consists of 57,000 digital objects (ca. 13GB). 97% of these objects are in 9 file formats, while the remaining 3% is in 100 different file formats.

with specific colour values attached to each pixel (see 2.2) they are more straightforward formats than vector images.

The only currently viable approach to preserving digital images is some form of format migration. As discussed above, emulation does not seem viable because of the need to continue to develop emulators as well as retain **all** original software applications. This is only one step better (i.e. no need to retain hardware) than the clearly unviable museum of technology approach. A more serious criticism of emulation has come from David Bearman who comments that "Rothenberg [the original and committed advocate of emulation] is fundamentally trying to preserve the wrong thing by preserving information systems functionality rather than records. As a consequence, the emulation solution would not preserve electronic records as evidence even if it could be made to work and is serious overkill for most electronic documents where preserving evidence is not a requirement".[83] Likewise, Bearman's criticism can be levelled at the UVC approach, which would have similar disadvantages for most institutions and organisations responsible for preservation.

Migration on request may offer another preservation option, but it is an approach whose worth can only be tested over time. As yet no operational preservation service has implemented such an approach. Until this approach has been implemented and tested operationally (as opposed to test situations), it is hard to see it as an approach that could be sustained by institutional repositories. There will be a sustained resource and programming burden in the development and maintenance of migration and rendering tools over time and this makes it unsuitable as an approach except in some specific circumstances (that probably will not reflect the reality of digital preservation situations).

As recommended in Chapter 5, the use of uncompressed baseline TIFF version 6 is currently the best strategy for preservation of raster images. JPEG2000 is also recommended as an acceptable preservation format with the qualification that it does not yet enjoy widespread industry support and in the long run may fail to gain enough acceptance to justify its recommendation as a preservation format. Images in JPEG2000 format can be retained in this format as preservation versions, but the qualifications attached to the format suggest that digital images in other formats should not be migrated into JPEG2000 for preservation purposes.

None of the other raster image formats are currently recommended for preservation, although there is the possibility that both PNG and DNG could become recommended formats in the future.

6.5 Vector Images

These are essentially a series of XYZ coordinates which define an image, although other information can be present in a vector image file. In some ways vector images can be regarded as less complex than raster images, with significantly smaller file sizes. However, vector images present more difficulties for preservation than raster images for a number of reasons. Firstly, there is a very large number of vector image software applications in use, with a consequent multitude of vector image file formats

---

[83] D. Bearman, "Reality and Chimeras in the Preservation of Electronic Records", *D-Lib Magazine*, vol. 5 no. 4 (April 1999). Available at: http://www.dlib.org/dlib/april99/bearman/04bearman.html. David Bearman is a seminal writer and thinker in the cultural heritage domain, specifically in the areas of archives/records and museums.

being created. Secondly, many (if not most) of the packages handling vector data use proprietary binary formats which are not appropriate for long term preservation. Thirdly, there are, as yet, no widely used non-proprietary open formats. Support within the GIS industry for the OGC data formats, although it is growing, is not yet widespread and they cannot be viewed as candidate preservation formats for GIS vector data.

Because there is such a wide range of uses of vector images, suitable preservation strategies will differ according to the use or purpose of the original vector data. Format migration may not be the best way to preserve all vector image files. There is, however, a generic 'lowest common denominator' approach to preservation of vector image data. This is to export the data as structured ASCII text. But note that without adequate documentation of the meaning of the text structure ASCII exports will be unusable. This extra requirement to explain the meaning of the ASCII text makes this approach more difficult to implement and adds to the cost and resource burden of the approach. Unfortunately, it is difficult to recommend any other approach as viable at present.

None of the approaches based on emulation or which rely on migration using specially developed and coded migration tools can be recommended as sustainable approaches at present. As discussed above, there is no practical experience or even pilot testing that would suggest any of these approaches as a viable method of preserving vector image files. The need for specialist programming skills and the software maintenance burden seem to work against the use of such approaches.

Clearly, there is a need for more research into the preservation of vector image files. This report cannot hope to come up with viable preservation strategies for file formats with such a range of factors which mitigate against easy preservation approaches. The preservation approaches proposed below are only interim solutions and should not necessarily be regarded as suitable for long-term preservation of such files. We recommend the initiation of appropriate research projects into vector image files as a high priority for JISC.

6.6 Recommendations for preservation approaches for digital images

6.6.i Raster/Bitmap Images

The recommended preservation formats for raster images are TIFF (uncompressed baseline version 6.0), and JPEG2000 with qualifications. The following table suggests possible approaches to preserving raster image files in the most widely used image formats.

Ingest Format (+ extension)	Notes
Tagged Image File Format (TIFF) .tiff; .tif	The best preservation format at this time (December 2005). Ensure it is uncompressed version 6.0, created without the use of additional application specific extensions ("baseline"). If so store unchanged. If compressed open with a suitable application (eg Adobe Photoshop) and save uncompressed.
JPEG2000 .jpg2; .jp2	Acceptable preservation format. Store unchanged.

Portable Network Graphics .png	Not suitable for preservation at this time (December 2005). Convert to the recommended TIFF format using an appropriate software application.
Bitmap .bmp	Not suitable for preservation. Convert to the recommended TIFF format using an appropriate software application.
Graphics Interchange Format .gif	Not suitable for preservation. Convert to the recommended TIFF format using an appropriate software application.
Joint Photographic Experts Group .jpg; .jpeg	Not suitable for preservation. Convert to the recommended TIFF format using an appropriate software application.
Digital Negative .dng	Not suitable for preservation at this time (December 2005). May become a suitable format for preservation of RAW image files produced by one-shot digital cameras. Convert to the recommended TIFF format using an appropriate software application.
Other raster image formats e.g. Corel Draw (.cdr), MrSid (.sid), Djvu (.djvu), etc.	Not suitable for preservation. Convert to the recommended TIFF format using an appropriate software application.

6.6.ii Vector Images

Because of the issues with vector image files (see above) it is not possible to recommend any preservation approaches as suitable for long-term preservation of vector image files. The approaches listed below are only interim suggestions and should not be relied upon for preservation of such files in excess of 10 years. ASCII formats encoded in XML (and variants) may offer the best hope for preservation. The following tables, categorised by software purpose, suggests possible approaches to preserving vector image files in most of the widely used formats.

General Vector Data applications

Format (+ file extension)	**Notes**
Scalable Vector Graphics .svg	Suitable preservation format for two-dimensional vector graphics (preferred).
Wavefront object files .obj	ASCII format suitable for preservation.
Computer Graphics Metafile .cgm	Suitable preservation format (a 2D format only).
Other vector image formats e.g. Macromedia Freehand (.af), CorelDraw (.cdr), Adobe Illustrator (.ai), etc.	Not suitable for preservation. Convert to .svg or ASCII.

Computer Aided Design (CAD) applications

Format (+ file extension)	**Notes**

Initial Graphics Exchange Specification (IGES) .iges	Suitable preservation format (preferred).
Standard for Exchange of Product data .step; .stp	Suitable format for preservation (preferred).
AutoCAD Drawing .dwg	Not suitable for preservation. Export from AutoCAD as STEP, or structured ASCII as a format of last resort.
AutoCAD Drawing Exchange Format .dxf	Not suitable for preservation. Export from AutoCAD as STEP, or structured ASCII as a format of last resort.
Other CAD formats e.g. FastCAD (.fcd), Bentley MicrosStation drawing format (.dgn); Helwett-Packard Graphics Language (HP-GL) (.hpl)	Not suitable for preservation. Convert to IGES or STEP.

Geographic Information Systems (GIS) applications

Format (+ file extension)	Notes
Geographic Markup Language .gml	An XML-based format suitable for preservation (preferred).
ESRI ArcInfo Export .e00	Suitable preservation format.
ESRI ArcInfo Ungenerate . poi; .lin; .pol; .txt	ASCII format suitable as preservation format of last resort.
MapInfo Interchaneg formats .mid; .mif	Suitable for preservation.
ESRI ArcView .shp; .shx; .dbf; .sbn; .sbx; .fbn; .fbx; .ain; .aih; .prj;	Not suitable for preservation. Convert to GML, or .e00.
MapInfo file formats .tab; .dat; .map	Not suitable for preservation. Convert to GML.
National Transfer Format .ntf	Not suitable for preservation. Convert to GML.
Map Overlay Statistical System (MOSS) .exp	Not suitable for preservation. Convert to GML.

Modelling and Animation applications

Format (+ file extension)	Notes
Virtual Reality Modelling Language (VRML) .vrml	Suitable preservation format since it is text-based.
Extensible 3D .x3d	An XML based extension of VRML. Suitable preservation format.

Macromedia Flash **.swf**	Not suitable for preservation. Existing conversion tools are problematic and only partially successful. Best approach may be to migrate to newer versions of Flash as necessary.
3D Studio Max .3ds	Not suitable for preservation. Export from application as IGES or ASCII.

Chemical, Biological, Crystallography applications

Format (+ file extension)	**Notes**
Crystallographic Information file .cif	Suitable preservation format.
Brookhaven Protein Databank file .pdbml	An XML based format suitable for preservation. Convert .pdb (ASCII based) to PDBML.
Elsevier MDL Molecule file .mol	Not suitable for preservation. Convert to CML.

7. Images Metadata Review and Requirements

7.1 Introduction

Metadata may be defined as structured information that describes resources. The effective use of archived digital images depends upon the suitability and adequacy of the metadata that describes them. This chapter aims to identify the metadata elements required to meet the needs of curators and end-users of archived digital images. It covers the essential technical, management and intellectual aspects of the image resource. It identifies the current metadata standards available for digital images and assesses two of the most recent – PREMIS, NISO Z39.87 – and their use within a METS environment. A usable, flexible and interoperable elements set for the archiving of digital images is recommended and areas requiring further work are identified.

Discussion about metadata is currently bedevilled by a confusion of terminology. For the sake of convenience here, metadata will be divided into three broad categories – technical metadata, management metadata and resource discovery metadata.

7.2 Technical metadata

Technical metadata describes the physical rather than intellectual characteristics of digital objects. There are relatively few *essential* technical metadata requirements for digital still images. The file format is the key detail that *must* be identified and recorded in order that the intellectual content can be decoded using appropriate software. The operating system and application software on the users' own machine will use this information to decide the application software appropriate to decode the image.

Fixity (authentication) information is essential in order for the curator of the digital image archive to be sure that an image has not been altered, either intentionally or unintentionally in any undocumented way. All digital objects, including image files, are mutable and are liable, even without external intervention, to "bit-rot" –the gradual degradation of stored bits leading to partial or even complete information loss". Fixity information should include the name of the fixity method, the value of any check-sum, and the date the fixity check was undertaken.

A small amount of image-specific information *must* also be recorded to identify the composition of the image itself and ensure the application software has interpreted it correctly. As discussed in Chapter Eight, a reliable measure of image quality - by which is meant the level of granularity, definition or detail contained within the image - is vital to both curators and end-users. The essential indicators are bit depth, colour space and image size or dimensions (measured by number of pixels in the horizontal dimension multiplied by the number of pixels in the vertical dimension). The resolution of the viewed or printed image may be inferred from the image size (total number of pixels) if the configuration of the output device is known. Different users will have very different ideas about what constitutes quality – according to the nature of their needs - but these three data elements provide the necessary information on which to base decisions about whether, in terms of its quality, an image is fit for the intended purpose. An array of other technical metadata can be recorded but these are the essential data elements.

7.3 Management metadata

Management metadata is information documenting the life cycle of an electronic resource, including data about ordering, acquisition, maintenance, licensing, rights, ownership, and provenance. It is essential that the provenance (custodial history) of a digital image object is recorded from, where possible, the time of its creation through all successive changes in custody or ownership. Users and curators must be provided with a sound basis for confidence that a digital image is exactly what it is purported to be. Likewise, all processing actions or changes that are made to the digital file(s) throughout the life cycle of the resource should be recorded and, in particular, any changes that result to the significant properties of the resource as a result. There should be a clear audit trail of all changes.

Any binding intellectual property rights attached to the digital image object must be documented in the metadata. Rights are crucial in two respects. They may limit the archive's powers to undertake preservation action on the resource. They may also prohibit or restrict its dissemination to users. A small amount of administrative metadata should also be created for each image, by which is meant metadata about the metadata-creation process itself. (i.e. the 'who', 'when' and possibly 'how' of its creation). As part of the 'business rules' or 'preservation planning' process of the archive, the curators will probably also wish to document their schedule for future preservation action in relation to the various classes of digital resource (e.g. next preservation action; date next preservation action due, retention periods, risk assessment, and significant properties of each class of image). However, it is not essential - nor perhaps desirable - for such information to be within the formal metadata of individual digital objects.

7.4 Discovery metadata

Technical and management *must* be augmented by resource discovery metadata. The purpose of archiving digital images is to facilitate their long-term retention *and use* in the future. Resource discovery metadata enables users to identify, locate and retrieve suitable images – normally via the medium of a searchable index or catalogue.

Intellectual content is a key criterion upon which resource discovery decisions are usually based. The content or subject or a digital image should be described so that the user can determine what is 'in' the image, what it is 'about' or what it represents. This is particularly important in relation to digital images because of the inherent lack of textual terms within an image object itself. The bare minimum is that each digital image has a unique permanent identifier but, wherever possible, suitable titles and free-text descriptions should be applied. Keywords from a controlled vocabulary may be added to enhance computerised retrieval. The latter is unlikely to be an automated process despite recent developments in image recognition software and is normally very labour-intensive.

Content is not, however, the sole determinant of suitability. Suitable images are those that are *fit for the purpose* the user has in mind. This brings us back to the importance of adequate technical and administrative metadata because suitability is related to the quality of the digital object - measured by resolution, bit depth and colour space – as well as its content.

7.5 Review of existing metadata standards

A review of metadata standards is timely because metadata provision has been identified as *the* chief issue causing concern within the archival image community. The section below starts with resource discovery standards and moves on towards management and technical standards.

7.5.i Dublin Core[84]

Simple Dublin Core is the best base standard upon which to develop a minimum generally applicable element set for the archiving of digital images. The *Dublin Core* standard (formally known as the Dublin Core Metadata Element Set) has defined fifteen core elements: Title, Creator, Subject, Description, Contributor, Publisher, Date, Type, Format, Identifier, Source, Language, Relation, Coverage and Rights. The worldwide DCMI consortium has designed the standard to accommodate the fundamental resource discovery requirements of *all* data types, subjects and domains. Because of its inherently cross-sector design it is well suited for use within a repository environment where digital images may be only one amongst a variety of data types. It is unlikely that repositories will find it feasible to implement separate schemas for each resource type.

Dublin Core is already widely adopted and is therefore a major source of interoperability between repositories. Simple DC metadata can be shared between all repositories, which support the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Widespread use of Dublin Core facilitates consistent results when users are searching across holdings in multiple repositories or browsing metadata records gathered from multiple repositories. All of the DC elements are relevant to digital images apart from language and, perhaps, publisher. All DC elements are optional and but a minimum element set for digital images should probably be restricted to Title, Creator, Subject, Date, Identifier, and Rights. Dublin Core is well suited to the description of digital images to which multiple rights may pertain because all of its elements – including Rights - are repeatable.

7.5.ii DC Terms[85]

Digital archives may legitimately prefer other metadata standards, such as DC Terms (previously known as Qualified Dublin Core) that accommodate fuller description than simple Dublin Core. DC Terms presents three elements that are not to be found in simple DC - audience, provenance and rightsHolder – and a set of element refinements (also called qualifiers) that refine the semantics of the elements and

---

84 http://dublincore.org/documents/usageguide/
85 http://dublincore.org/documents/usageguide/qualifiers.shtml

afford a greater degree of descriptive precision. Provenance, in particular, is a key piece of information for archives to record and it is recommended here that the dcterms:provenance element should be included in the minimum element set for archiving still digital images.

For the sake of interoperability and resource sharing, the key thing, though, is to choose a metadata standard that maps well to Dublin Core. Metadata in any format can be oai harvested, although only simple Dublin Core elements approach being universally interoperable. Digital image archives must judge for themselves (in relation to their particular circumstances) the optimal balance to strike between richness of description and maximum interoperability because they stand in opposition to one another.

7.5.iii VRA Core 3.0[86]

VRA Core 3.0, for instance, offers rich description of ‘works of visual culture’ to and is particularly suited to digital reproductions because a single record can describe a digital image (e.g. tiff image of a chair), the original analogue object (e.g. the chair), and the surrogate analogue (e.g. slide of the chair). It contains some elements of relevance to digital preservation, such as dimensions, format, resolution and rights, and is extensible. But it is not as interoperable as DC or DC Terms – being only suitable for the visual art community - and does not contain a comprehensive set of technical metadata elements.

Digital reproductions are usually more complicated to describe than ‘born digital’ images because there is a source entity to consider. The Dublin Core record should focus unambiguously on the digital object. DC:date, for instance, should be the date the digital image file was created or data captured rather than the date of the source image (e.g. the painting). DC:source, dc:relation, and dc:description may provide important information about the source item, but the orientation as a whole should be towards the digital object.

7.5.iv PREMIS[87]

Dublin Core is designed for purposes of resource discovery and lacks technical and management metadata elements. Therefore, the minimal Dublin Core element set should therefore probably be augmented by elements from the PREMIS data dictionary[88]. PREMIS defines a set of core management and technical metadata elements that are needed to support the preservation of all kinds of digital resources, regardless of their data type (i.e. not just digital images). It covers the entities *Objects, Events, Agents, and Rights.* Objects includes all of the essential technical metadata requirements of digital still images that are of a generic sort (i.e. applicable to all data types) - such as fixity, format name and format version. Rights covers all of the rights and permissions required for preservation activity but, as the PREMIS working group itself acknowledges, the focus is specific so that rights associated with access and/or dissemination are out of scope - as are all resource discovery requirements on the basis that they are provided for by other standards such as Dublin Core. PREMIS is intentionally not a comprehensive standard that meets all

86 http://www.vraweb.org/vracore3.htm
87 See *Preservation Metadata: Implementation Strategies (PREMIS)* international working group home page at: http://www.loc.gov/standards/premis/
88 http://www.oclc.org/research/projects/pmwg/premis-final.pdf

the metadata needs of a digital repository but should provide most of the essential technical and management metadata elements necessary for the archiving of digital still images.

7.5.v NISO Z39.87[89]

PREMIS adopts a generic approach similar to Dublin Core in that it provides a core set of technical metadata elements for describing the things that are common to all digital formats. For format-specific detail, such as colour space and bit depth, repositories should augment it with elements from more granular standards. For raster images, that should probably mean the draft NISO Z39.87 standard: *Data Dictionary – Technical Metadata for Digital Still Images.* Z39.87 has an XML schema (MIX) [90] which is an extension of the METS schema and stands every chance of widespread uptake and long-term support because it has the support of OCLC, RLG and the Library of Congress. It is supported – albeit patchily - by technical metadata extraction tools such as JHOVE.

Z39.87 was designed to cover *all* aspects of technical metadata required for the management and preservation of raster. The basic digital object information is about the rendering of a viewable image; image capture metadata describes the analogue-to-digital conversion process; image assessment metadata provides “anchors” to assess the accuracy of the digital image output; and there is change history metadata. The specificity of Z39.87 makes it obviously well suited to the raster formats for which it is designed - some elements even relate to specific formats such as JPEG2000 or MrSID. This strength, though, may also be its biggest weakness. A strategy whereby repositories invoke a separate standard for each resource type or format that they archive is unlikely to be sustainable. For the sake of interoperability, archives should be encouraged to adopt generic standards wherever possible. Fortunately there is good convergence between the generic elements of PREMIS and the mostly-mandatory basic file information section in Z39.87 and a good approach would be to avoid, as far as possible, the use of other, non-mandatory Z39.87 elements.

7.6 Standards for vector images

Vector graphic formats are diverse and are used for many purposes. Consequently, recommending a single element set for vector graphics is more of a problem than for raster images. Several standards exist that, although widely used in particular fields of expertise, have little use in the wider community.

7.6.i CAD

Metadata can be stored in CAD files and it is customary in architectural drawings to include blocks in the border drawing of the drawing that document the contents. There are ISO standards for the contents of these blocks (ISO 7200). There are various standards (e.g. ISO 13567) for naming layers in CAD drawings which generally attempt to make these self-documented. The layer name indicates some

---

[89] http://www.niso.org/standards/standard_detail.cfm?std_id=731
[90] http://www.loc.gov/standards/mix/

common property of its contents. STEP and IGES aim to contain full and extensive metadata.

However, there is no general metadata schema for CAD files. The Centre for the Study of Architecture/Archaeology specifies the required documentation to accompany deposits to its CAD archive[91] and the section on documentation of CAD models in *CAD: A Guide to Good Practice*[92] provide the nearest there is to suitable metadata standards. For architectural drawings the Categories for the Description of Works of Art (CDWA)[93] is considered suitable for resource discovery purposes although is overly detailed for many architectural drawings, which cover such mundane subject material as plumbing diagrams, drainage plans etc.

For other (i.e. non-architectural) uses of CAD, the nature of the metadata may be dependant on the purpose of the drawing/model. Generically, metadata is often taken to be descriptions of the non-visual elements of a drawing - layers; blocks; text style definitions; object attributes; links to databases, images, URLs; print definitions and such like. These are useful/essential for the re-use of a CAD drawing and should be included in the metadata. Otherwise marginal blocks appear to be the only standard for including metadata (ISO/TR 19033:2000) and this is also the case for engineering drawings. The concept remains that these are really paper drawings and have only incidentally been created on computers.

The users of CAD for geospatial purposes should be aware that CAD drawings do, or can, contain more (non-drawing) elements/information than standard GIS formats and that the non-drawing elements also need to be included along with the standard GIS metadata.

7.6.ii GIS

ISO 19115[94] (and ISO 19139, the XML schema implementation, so far incomplete) is widely considered to be the ultimate GIS metadata standard and must therefore be the recommended standard. UK GEMINI[95]**,** the new UK national geo-spatial metadata profile, is compliant with ISO 19115. UK GEMINI allows for the creation of discovery metadata with both ISO 19115 (*Geographic Information –Metadata*) and the national *e-Government Metadata Standard* (eGMS), ensuring compliance with both.

The NZ Geospatial Metadata Specification)[96] is a subset of ISO 19115 and therefore compliant in so far as it contains all of the mandatory elements of it and selected optional extensions particularly suited to New Zealand needs. Work is being carried out to bring other major GIS metadata standards, such as the Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata (CSDGM)[97],

---

91 http://csanet.org/archive/archivedoc.html
92 http://ads.ahds.ac.uk/project/goodguides/cad/sect56.html

93 www.getty.edu/research/conducting_research/standards/cdwa/

94http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26020&ICS1=35
95 www.gigateway.co.uk/metadata/pdf/UK_GEMINI_v1.pdf , (Geo-spatial Metadata Interoperability Initiative)

96http://www.linz.govt.nz/rcs/linz/pub/web/root/core/Topography/ProjectsAndProgrammes/geospatialmetadata/index.jsp

97 www.fgdc.gov/metadata/meta_stand.html which has a Remote Sensing Extension[98], a Biological Data Profile[99], and a Shoreline Data Profile[100], into compliance.

CML (Chemical Markup Language) appears set to become the standard format for chemical data. The core elements (the actual molecular description) are complete and published. A number of other elements are still under development[101]. However there appears to be no description or preservation metadata standards for chemical data.

7.6.iii VRML in X3D

VRML is Virtual Reality Modelling Language for 3D computer-generated graphics and 3D sound. In VRML in X3D, the latest revision of the VRML standard, the head section of a file may contain html-style meta tags in which any name value pairs can be used to store metadata.

7.6.iv Scalable Vector Graphics (SVG)[102]

SVG is a text-based markup language for describing object layout that uses coordinates to indicate the size, shape and fill of an object. As an XML schema, it is possible to import elements from other namespaces in order to describe the resource itself.

7.7 Collection-level metadata for image collections

Resource discovery metadata should, where possible, be created at collection, as well as item, level. Collection-level records enhance resource discovery and provide a home for contextual information about the wider environment within which the digital image was created – for instance, the name, purpose and nature of the parent project, any grant number associated with it, and the other images related. Such information gives important clues for interpreting the image and both from a logical and practical point of view, is best recorded once and referenced by each item-level record in the collection.

7.8 METS[103]

METS can act as a suitably flexible container or wrapper for image metadata of all kinds. Metadata expressed using the various schemas above can be held within separate sections of a METS wrapper and linked to the relevant data file. The METS

98 http://www.fgdc.gov/standards/status/csdgm_rs_ex.html
99 http://www.fgdc.gov/standards/status/sub5_2.html
100 http://www.fgdc.gov/standards/status/sub5_6.html
101 http://www.xml-cml.org/
102 http://www.w3.org/TR/SVG/metadata.html#MetadataElement
103 http://www.loc.gov/standards/mets/

structural map will, for instance, associate each MIX element with the correct data file from among the many that may be generated during the lifecycle of the image. METS is fast emerging as the metadata wrapper of choice for digital repositories. The METS document, in turn, is an instantiation of an Information Package as stipulated in the Open Archival Information System (OAIS).

7.9 Contextual documentation

The key is to implement a suitable metadata schema within an OAIS compliant structure. However, there may be occasions where additional, less structured, contextual information should be collected that describes the process and methodology by which the images have been created or captured. Vector images are generated from structured data and as such they generally have a single table or file associated with them containing XYZ values. For structured data exported as delimited text from static databases (a long-term preservation strategy) it is considered good archival practice to create an Entity Relation Model (ERM) representing structure and a data dictionary describing the attributes within each entity and to maintain these as part of the archival package. An alternative would be to export as XML using a schema representing structure and attributes, which remove the necessity of storing associated files - although this significantly increases the complexity of creating the archival package because individual schemas have to be created[104].

In theory, at least, similar archiving techniques could be applied to structured data associated with vector images; however, the simplicity of the data structure suggests its essence could be maintained in associated metadata. In any case packages generating vector images are increasingly using XML schemas to store such data.

7.10 Metadata extraction and storage

7.10.i Creating metadata

The extraction of technical metadata manually from image files is largely impractical. A small number of tools have therefore been developed to do it automatically, although there is little available for vector images. The uptake of specific metadata standards, such as PREMIS and Z39.87, is likely to be influenced in part by the quality of the programmes designed for them. The majority of data elements stipulated by NISO Z39.87 could be recorded automatically during image capture, at least in theory. Z39.87 has been designed so that technical metadata embedded in the headers of TIFF Rev. 6.0, TIFF/EP, and EXIF file formats, as well as some metadata elements embodied within the just-emerging Digital Imaging Group's DIG35 metadata element set will map to it. The main challenge will be to persuade camera manufacturers to record this data and, as the RLG *Automatic Exposure*[105] survey concluded, the record, to date, is patchy.

The following are freely available metadata harvesting tools.

---

[104] See http://www.erpanet.org/events/2003/bern/Bern_Report_final.pdf. For a fairly recent discussion of preservation techniques for databases
[105] http://www.rlg.org/longterm/ae_whitepaper_2003.pdf

7.10.ii JHOVE[106]

JHOVE is the most useful metadata extraction tool for many purposes but it still only supports jpeg, jpeg2000, gif and tiff image formats. Future development work should be undertaken to extend the range of JHOVE modules so that it can identify other formats (e.g. png, bmp). Such modules should open source. The quality and amount of metadata extracted from images using JHOVE varies depending on the file type. For example the gif output includes all the information contained in the Header, Logical Screen Descriptor, Global Colour Table, Image Descriptor, Local Colour Table and Image Descriptor sections of the file whereas the tiff output will recognise a different set of properties specific to that format. JHOVE can output in a simple XML schema (jhove.xsd) in which 'name' and 'value' pairs are children of 'property' tags allowing a great deal of flexibility but little real structure. There is no official namespace. Output is to a variety of MIX and JHOVE-specific elements, with the latter sometimes being preferred even when a suitable MIX element exists. (e.g. JHove has <size> and Mix has <fileSize>) It is recommended that JHOVE and other metadata extraction tools should generate metadata that is compliant with PREMIS and the mix schema.

7.10.iii National Library of New Zealand Metadata Extract tool[107]

The National Library of New Zealand Metadata Extract tool supports Windows bitmap, gif, jpeg, and tiff formats. There is a single basic metadata schema with a series of extensions based on the file type. The whole is more consistent but less detailed, than the output produced by JHOVE. It extracts a very limited element set. There are some implementation problems, for example the tool fails to report the compression of jpeg images correctly. It can also be fooled because it identifies the file format on the basis of the file extension. Output from the program is in National Library Preservation Metadata Data Dictionary XML although the documentation suggests this is user configurable. The National Library of New Zealand Metadata Extract tool is open source like JHOVE but there has been less take-up and the National Library has not committed to institutional support for it. However, unlike JHOVE, it can handle complex relationships – for instance defining website files and relationships between them or spreadsheets.

There are also commercial packages that extract metadata from CAD files but they are generally rather vague in defining what constitutes metadata.

Unlike JHOVE, it can handle complex relationships – for instance defining website files and relationships between them or spreadsheets.

7.11 Storing Metadata

Essentially there are two, or perhaps three, ways of storing image metadata. Metadata may be embedded in the digital objects, be held separately as a distinct metadata record, or both. Traditionally metadata has been stored in a database

---

106 http://hul.harvard.edu/jhove/
107 http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#extraction

separate from the image itself. This is fine when the image remains with its data in, say, an online catalogue. But it does present problems when the image is downloaded and removed from its original setting.

Embedding information in the image file itself can be a solution. Just about every image file-type that we use *can* hold this tagged information, the problem is in creating a standard way for it to be read and written by the image-readers and image-editing software. The two common standard formats used in the imaging/photographic industry for embedding file information within an image are: International Press Telecommunications Council (IPTC) [108] and EXchangable Image File (EXIF) [109]. EXIF metadata automatically generated by digital cameras and stored in the image files may include shutter speed, date taken, aperture, GPS information, and other information depending on the make of the camera. EXIF data cannot be modified, only viewed or copied into new images. Another common format, Photoshop's File Info, is based on a subset of IPTC.

We would not recommend relying on embedded information alone. There remain various issues regarding the use of the common formats and their ability to be read in common image application software. For example older versions of Photoshop (pre v. 7) can read IPCT information via its File Info function, but only v. 7 and later can read EXIF formatted tags. However the editing of embedded tags is possible in Image Management Systems such as Canto Cumulus (IPTC), Extensis Portfolio (EXIF/IPTC), Thumbsplus (IPTC), FotoStation (IPTC) and iView (EXIF/IPTC). Imatch (EXIF/IPTC) also allows for bulk editing of information and has the ability to add new user defined fields.

7.11.i Embedding metadata in PNG[110]

The PNG's file structure is fairly simple. After identifying itself as a PNG image, it is organised into a series of 'chunks' beginning with an IHDR (image header), ending with an IEND, and with a number of IDAT (image data) chunks in between. It may optionally have chunks defining the palette, gamma and colour profile, and text chunks, enabling basic description and keywords (several are predefined, such as title, author, copyright, disclaimer and source). It is also possible to register new chunks of textual information, offering the possibility of providing more structured metadata in the future.

7.11.ii Embedding metadata in JPEG2000[111]

JPEG 2000 Part 1 offers simple, unstructured support for metadata, enabling those creating JPEG 2000 images to add blocks of text to the .JP2 file. Structured metadata is specified for the extended JPEG 2000 Part 2 images (JPX), although TASI has not yet found an implementation of this. The JPX metadata categories were developed specifically for JPEG 2000, but are based on the DIG35 standard sponsored by the Digital Imaging Group (now the I3A)[112].

---

108 http://www.iptc.org/pages/index.php
109 See www.EXIF.org
110 http://www.libpng.org/pub/png/
111 http://www.jpeg.org/jpeg2000/metadata.html
112 http://www.i3a.org/i_dig35.html

In short, with its XML construct, a single JPEG 2000 file may contain technical information about the image creation process (e.g. camera/lens specification and capture conditions); change history metadata about the processing steps applied (digital capture, exposure of negative or reversal films, creation of prints, scans of negative or positive film or of prints); resource discovery elements into which users can enter keywords or phrases covering the "who", "what", "when" and "where" aspects of the image; and also IPR/copyright details, provenance, permissions and perhaps even security devices. Mappings are being made to other metadata standards, such as the Dublin Core.

Projects are currently underway working in this area. For instance, Archives & Special Collections at the University of Connecticut have stated in an email:
"One of the many interesting features of this new standard is the ability to embed XML boxes within the image file. We have successfully embedded in a single (albeit rich) file the following elements: a lossless compressed image (no loss of data from TIFF to JPEG2000 but greatly reduced in size, with the ability to display low to high resolutions without creating additional image files); automatically derived technical metadata in XML; descriptive metadata in XML; PDF files; and entire EAD finding aids. While we are just beginning to explore the potential uses for this capability, we are excited that, for digitised individual items from a collection, we are able to embed within these digital objects their contexts in the form of EAD finding aids. So, no matter where we use, move, or combine these individual digital objects, their contextual metadata (EAD finding aids, full or brief versions) move with them."

7.11.iii Embedding metadata in TIFF[113]

The TIFF specification defines a framework for an image header called 'IFD' (Image File Directory) that is essentially a flexible set of specifically those tags that the TIFF writer software wishes to specify. The clear benefit of this scheme is that almost any information can accompany an image, while little information is absolutely needed, and image headers remain as lean as possible. There is little overhead, and enough flexibility to suit any need. Tiff allows the inclusion of an unlimited amount of private or special purpose information[114]. It may be that TIFF 6.0 does allow a separate EXIF IFD, although that is not entirely clear[115].

7.11.iv Embedding metadata in DNG[116]

The DNG format is fully able to read all image tagging (EXIF & IPTC) from the original RAW file and transfer it to the new DNG image. It is a non-proprietary format defined by Adobe as a solution to the problem of manufacturers dropping support for proprietary raw formats shortly after the cameras that use them are discontinued - with the danger that images created in those formats may be lost forever. DNG is an extension of the TIFF 6.0 format, and is compatible with the TIFF-EP standard. Its metadata standard is publicly documented which means that software readers such as the Adobe Camera Raw plug-in do not need camera-specific knowledge to decode and process files created by a camera that supports DNG. Additional metadata may be embedded in DNG using TIFF-EP or EXIF metadata tags (which

[113] For a list of baseline tags see: http://www.awaresystems.be/imaging/tiff/tifftags/baseline.html
[114] http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf
[115] http://www.asmail.be/msg0055176682.html
[116] http://www.adobe.com/products/dng/pdfs/dng_spec.pdf

use nearly the same metadata tag set as each other), the IPTC metadata tag (33723) or the XMP metadata tag (700).

Much of the technical metadata of raster image files is integral to the format and some formats, e.g. png[117] and JPEG2000, can also contain RDF style metadata. Adobe has developed Extensible Metadata Platform (XMP)[118] to embed metadata into files but presumably software rendering these files must be modified to accept the extra data. It is also possible to embed the image file in an XML document containing descriptive metadata using tools such as Xena[119] from the National Archives of Australia.

7.12 Conclusion

Although it is desirable to encapsulate objects, storing both the metadata and the digital object a single structure, metadata required to manage digital archives needs to be accessible and maintainable. Therefore extracting and storing all the relevant metadata in a separate file is preferable to having to access the objects themselves to retrieve metadata.

While the cultural heritage community has defined standard metadata element sets (e.g. PREMIS and Z39.87) for the digital preservation of still images, the camera manufacturing industry has launched a number of initiatives that promise to deliver self-describing digital files, or files that carry within their code vital information about their origination, content, access rights, etc. In some instances, these initiatives propose metadata element sets that include tags relevant to digital preservation (such as DIG35 or EXIF); in other instances, they propose specific or generic transfer mechanisms for self-describing metadata (such as the XML box in JPEG 2000's JP2 and JPX file types or **Adobe's eXtensible Metadata Platform, Adobe XMP**). The industry at large and the manufacturers of digital capture devices have already made an investment by developing and implementing some of these technologies. But to date there are no tools available for automatically extracting the whole range of elements in Z39.87[120]. Z39.87 attempts to provide an ideal, comprehensive, array of elements but the reality is the only file level metadata that is likely to exist is that which can be auto-generated. Auto-generated file-level metadata is going to minimal for most images but possibly more extensive for born digital images (EXIF, etc).

[117] http://www.tasi.ac.uk/2000/09/rdfmeta/
[118] http://www.adobe.com/products/xmp/main.html
[119] http://xena.sourceforge.net/
[120] Dale, R.L., & Waibel, S. (2004). *Capturing Technical Metadata for Digital Still Images.* RLG DigiNews, Oct 15, 2004. http://www.rlg.org/en/page.php?Page_ID=20462&Printable=1&Article_ID=1676

7.13 A recommended minimum element set for use with archival digital still images

The proposed minimum element set for digital still images contains elements from:

- Simple Dublin Core - resource discovery elements
- PREMIS - generic management and technical metadata
- A small number of elements drawn from NISO Z39.87 to be added to this core when describing raster images.
- There is no single standard for describing vector graphics and, in view of the diverse range of file formats and their many uses, it is not feasible here to list a definitive element set for them.

No.	Element	Metadata standard	Definition	Comment
1	Title	Dublin Core	The name given to the resource.	Typically, a Title will be a name by which the resource is formally known.
2	Creator	Dublin Core	An entity primarily responsible for making the content of the resource.	Examples of a Creator include a person, an organization, or a service. Typically the name of the Creator should be used to indicate the entity.
3	Subject	Dublin Core	The topic of the content of the resource.	Typically, a Subject will be expressed as keywords or key phrases or classification codes that describe the topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or

				formal classification scheme.
4	Date	Dublin Core	A date associated with an event in the life cycle of the resource. format.	Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [Date and Time Formats, W3C Note, **http://www.w3.org/TR/NOTE-datetime**] and follows the YYYY-MM-DD
5	Identifier	Dublin Core	An unambiguous reference to the resource within a given context.	Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Examples of formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL), the Digital Object Identifier (DOI)
6	Provenance	Dublin Core	A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity and	The statement may include a description of any changes successive custodians made to the resource.

			interpretation.	
7	Rights	Dublin Core	Information about rights held in and over the resource.	Typically a Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the rights element is absent, no assumptions can be made about the status of these and other rights with respect to the resource.
8	objectIdentiferType	PREMIS	A designation of the domain within which the object identifier is unique.	Identifier values cannot be assumed to be unique across domains; the combination of objectIdentifierType and ojectIdentifierValue should ensure uniqueness. Value should be taken from a controlled vocabulary.
9	objectIdentiferValue	PREMIS	A designation used to uniquely identify the object within the preservation repository system in which it is stored.	The value of the objectIdentifier.

10	preservationLevel	PREMIS	A value indicating the set of preservation functions expected to be applied to the object.	Some preservation repositories will offer multiple preservation options depending on factors such as the value or uniqueness of the material, the "preservability" of the format, the amount the customer is willing to pay, etc. Value should be taken from a controlled vocabulary.
11	ObjectCategory	PREMIS	The category of object to which the metadata applies.	Preservation repositories are likely to treat different categories of objects (representations, files, and bitstreams) differently in terms of metadata and data management functions. Value should be taken from a controlled vocabulary.
12	messageDigestAlgorithm	PREMIS	The specific algorithm used to construct the message digest for the digital object.	Value should be taken from a controlled vocabulary.
13	messageDigest	PREMIS	The output of the message digest algorithm.	This must be stored so that it can be compared in future fixity checks.
14	formatName	PREMIS	A designation of the format of the file or bitstream.	Value should be taken from a controlled vocabulary.
15	formatVersion	PREMIS	The version of the format	Many authority lists of format names are not granular

			named in formatName.	enough to indicate version, for example, MIME Media types.
16	storageMedium	PREMIS	The physical medium on which the object is stored (e.g., magnetic tape, hard disk, CD-ROM, DVD).	The repository needs to know the medium on which an object is stored in order to know how and when to do media refreshment and media migration.
17	eventIdentifierType	PREMIS	A designation of the domain within which the event identifier is unique.	For most preservation repositories, the eventIdentifierType will be their own internal numbering system. It can be implicit within the system and provided explicitly only if the data is exported.
18	eventIdentifierValue	PREMIS	The value of the eventIdentifier.	
19	eventType	PREMIS	A categorization of the nature of the event.	Categorizing events will aid the preservation repository in machine processing of event information, particularly in reporting. Value should be taken from a controlled vocabulary.
20	eventDateTime	PREMIS	The single date and time, or date and time range, at or during which the event	Any date/time convention may be used, as long as it is consistent and can be translated into ISO 8601 for export if necessary.

			occurred.	
21	agentIdentifierType	PREMIS	A designation of the domain in which the agent identifier is unique.	Value should be taken from a controlled vocabulary.
22	agentIdentifierValue	PREMIS	The value of the agentIdentifier.	May be a unique key or a controlled textual form of name.
Essential additional elements for raster images				
23	colorSpace	NISO Z39. 87	A designation of the colour model of the decompressed image data.	Commonly called colour spaces, these colour models (e.g. RGB, YcbCr) are drawn from common file formats used to render digital still images. Some colour models may be pertinent to certain files types (e.g., TIFF) while others are more device dependent or independent (calibrated) colour models. colorSpace should be a text description.
24	imageWidth	NISO Z39.87	A specification of the width of the digital image, i.e.horizontal or X dimension, pixels.	The image width may be the shorter or longer dimension of the image, depending upon the orientation of the camera or scanner during image capture. For multiple-resolution image file formats, value shall specify the highest resolution.
25	imageHeight	NISO Z39.87	A specification of the height of the digital	The image height may be the shorter or longer dimension of the image, depending upon

			image, i.e. vertical or Y dimension, in pixels.	the orientation of the camera or scanner during image capture. For multiple-resolution image file formats, value shall specify the highest resolution.
26	bitsPerSample	NISO Z39.87	The number of bits per component for each pixel.	This field is used to describe the number of bits for each sample (or channel), expressed in the same order given in colorSpace. BitsPerSample is equivalent to bit depth. It gives the sample rate per colour channel – so, for instance, 8,8,8 is 24bit.

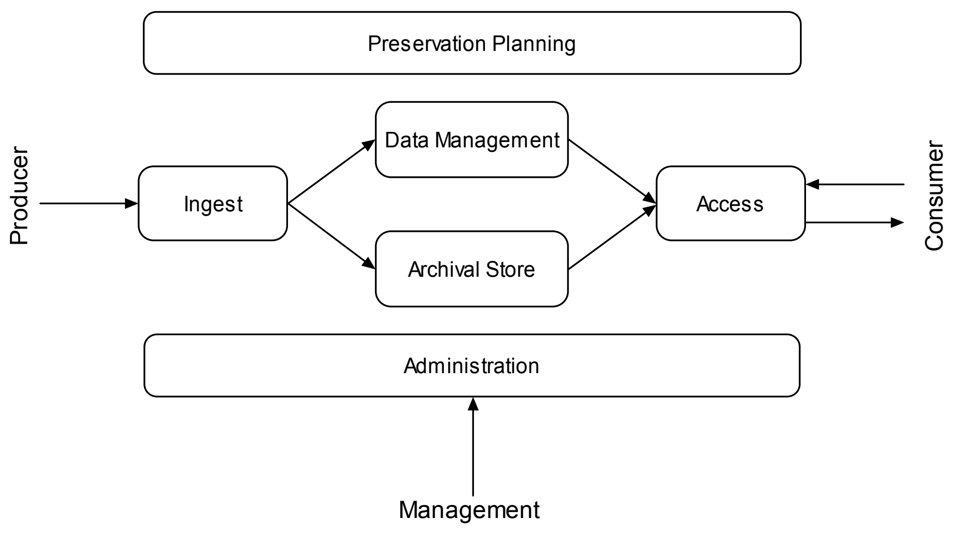
8. Life Cycle and Organisational Models

A strategic approach to managing the life cycle of digital collections has been broadly advocated and a significant amount of work has been undertaken over the last 10 years or so to identify and map the life cycle of digital objects. The concept of the life-cycle is a useful tool that allows us to identify the key events that take place in the life of a digital object, to identify the actions that we need to take at these key events, and the supporting policies and documentation that are needed to underpin these events and actions. It allows us to plan for the long-term preservation of digital objects in a constructive and time-based fashion, reflecting the more active, interventionist methods of preservation that are required for digital objects, and it allows us to construct organisational models that can support these activities.

Before presenting the life-cycle and organisational models developed for images it is useful to have a more detailed understanding of what the functional requirements of an image preservation repository – one that is capable of preserving digital images in the long-term – would be, and what attributes it would need to possess to be regarded as a secure place of deposit for long-term preservation of digital images to inform our thinking.

8.1 Functional Requirements: OAIS

The functional requirements for the preservation of digital information have been the focus of considerable attention and the *Reference Model for an Open Archival Information System (OAIS)* (Consultative Committee for Space Data Systems [CCSDS], 2002) has become the accepted standard.



8.2 OAIS Functional Entities (Simplified)

Source: Based on Figure 4-1 in CCSDS, 2002, p. 4-1

The OAIS functional model, shown above, identifies the main tasks that any type of repository must perform in order to secure the long-term preservation of digital material. The model defines six main functional entities that describe the activity of a digital repository as a flow of digital material, from the arrival of new material in the repository, its storage and management, and through to its delivery to a user (consumer).

Ingest

Ingest includes the physical transfer of files and the legal transfer of rights through the signing of licences or other agreements that establish the OAIS repository's right to maintain the ingested material. During ingest, descriptive information (resource discovery metadata) should be created to describe the material, and the submitted files are checked to ensure that they are consistent with the OAIS repository's data formatting and documentation standards. This may include tasks such as file format conversions or other changes to the technical representation and organisation of the submitted material.

Archival Storage

This functional entity is concerned with the bit storage of the submitted digital material including tasks such as backup, mirroring, security and disaster recovery.

Access

All the services and functions needed for users to find and access the contents of the repository.

Data Management

Data management involves the collection, management and retrieval of both resource discovery, administrative and preservation metadata.

Administration

The administration functional entity involves the entire range of administrative activities that an archival organisation should undertake. Notable tasks include managing, monitoring and developing the repository's software systems, negotiating submission agreements with producers (authors), and the establishment of policies and standards for the repository.

Preservation Planning

This functional includes four sub-entities associated with identifying preservation risks and developing plans to address them:

Monitor Designated Community – the designated community is an OAIS term that refers to the community of stakeholders who have an interest in the content of the repository. An OAIS repository needs to monitor its designated community's adoption of new technology, and other trends that may affect preservation of the community's digital output. In the case of digital images, this would refer in part to the user communities identified in Chapter Three.

Monitor Technology – The monitor technology function ensures that the OAIS repository is constantly aware of technological changes that may render its current holdings obsolete or difficult to access.

Develop Preservation Strategies and Standards – The development of strategies and standards for preservation that are informed by the current and future requirements of the producers and consumers of the OAIS repository.

Develop Packaging Designs and Migration Plans – This function accepts standards for file formats, metadata and documentation (generated as part of the administration functional entity) and creates tools or defines techniques that apply these standards to submissions.

8.3 Trusted Digital Repositories (TDR)

Over recent years work has been carried first by RLG and OCLC, and subsequently by RLG and NARA (further information can be found at: http://www.rlg.org/), to develop a model and checklist for the attributes of a trusted digital repository. The concept of TDR is intended to provide an environment of trust between content owners and creators and those who are be responsible for its long-term preservation, similar to that which exists between authors, publishers and libraries. A trusted digital repository is one whose mission is to provide reliable, long-term access to managed resources now and in the future. The organisational structure behind such a trusted digital repository is regarded as flexible. What is important is the ability to meet expectations and to be able to demonstrate the following attributes:

Organisation

- i. governance and organisational viability: a repository must demonstrate an explicit, tangible, and long-term commitment to compliance with prevailing standards, policies and practices
- ii. organisational structure and staffing: a repository must have designated staff with requisite skills and training and must provide ongoing development
- iii. procedural accountability and policy framework: a repository must provide clear and explicit documentation of its requirements, decisions, development and action to ensure long-term access to digital content in its care
- iv. financial sustainability: a TDR should be able to prove its financial sustainability over time
- v. contracts, licences and liabilities: a repository has and maintains appropriate contracts and deposit agreements

Repository Functions, Processes and Procedures

These will differ between repositories but the key issues are that the policies, procedures, functions and processes are properly documented and available for public scrutiny, and that the repository is following best practice as outlined in the OAIS functional model and as specified for the content types with which it is working.

- i. ingest/acquisition of content
- ii. archival storage and management of archived information
- iii. preservation planning, migration, and other strategies
- iv. data management
- v. access management

The Designated User Community and the Usability of Information

- i. documentation: the repository has a definition of its designated user community/ies and what levels of service it expects, and that this is a public document

ii. descriptive metadata: a repository must articulate minimum metadata requirements to enable the user community to discover and identify materials of interest
iii. use and usability: access and delivery options are open and transparent and fit for purpose, and that ensure that all legal aspects are complied with
iv. verifying understandability: to have mechanisms in place that ensure the repository obtains, manages, and makes available information in forms that allow digital objects to be understandable and usable over time

Technologies and Technical Infrastructure

i. system infrastructure: the repository must provide a secure and trusted infrastructure to manage its digital content
ii. appropriate technologies: a repository should use strategies and standards relevant to its designated communities and its digital technologies
iii. security: the system must be secure and protected, including m-m interaction and human – machine interaction

8.4 Digital Image Life-cycle Model: OAIS and TDR

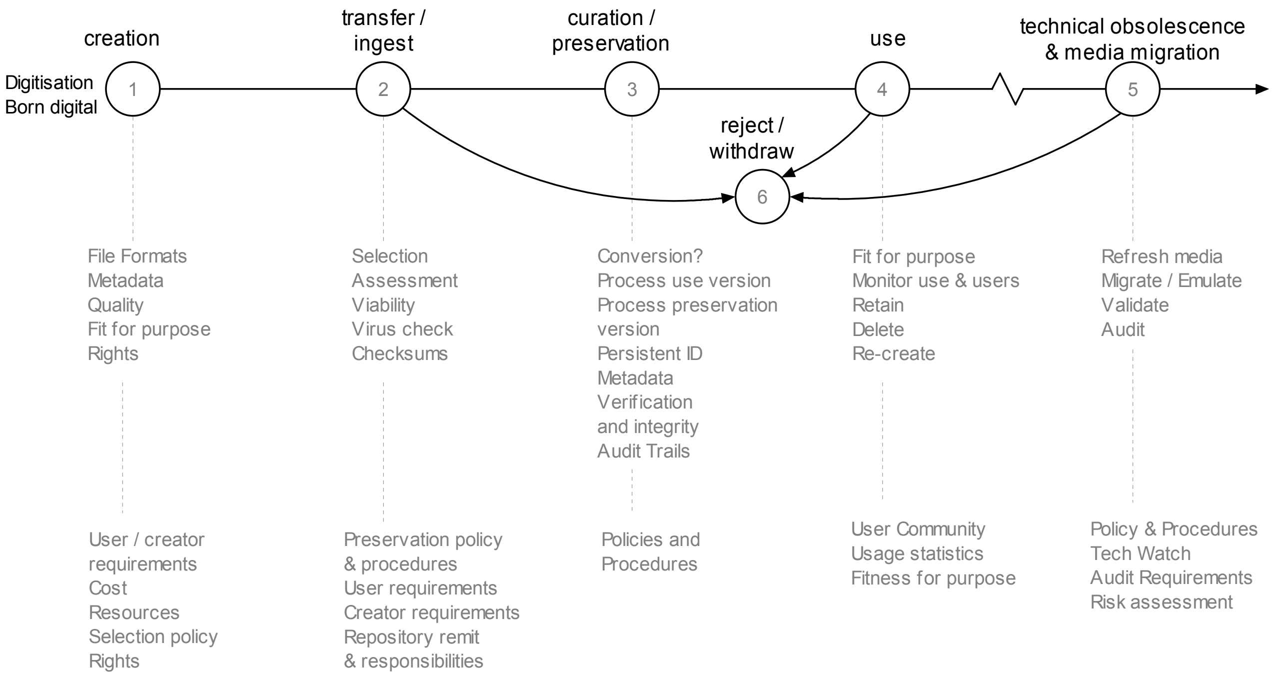
The OAIS model offers a functional model for any digital preservation system. OAIS itself does not provide a practical implementation of the framework however, and institutions will adopt their own locally appropriate workflow solutions. Similarly, the Trusted Digital Repositories framework provides a checklist of things a repository should do if it wishes to be taken seriously in its preservation efforts, but does not specify how that might relate to workflows.

The life-cycle model presented here attempts to incorporate a sense, in a somewhat simplified schematic model, of how the life-cycle meets OAIS and TDR. It identifies the key events that take place, the activities that should take place at those events, and the policies and processes which underpin them. Thus the top layer conveys the idea that this is part of continuum where key actions points are identified; the second layer outlines the actions and decisions that are likely to made at this time; and the third layer identifies the requirements, policies and processes on which these decisions are likely to be based. Readers will hopefully recognise elements from both OAIS and the TDR.

What follows is intended to bring together much of what has gone before in the previous chapters of this report by highlighting the key issues, challenges and decisions to be taken at key points in the digital image life-cycle. It is also intended to provide practical and useful advice for those responsible for the preservation and curation of digital image collections. The model presents six key events that may occur in the full lifecycle of digital images:

1. Creation
2. Transfer/Ingest
3. Curation/Preservation
4. Access and Use
5. Technical Obsolescence and media migration
6. Withdraw / reject

At each key event a range of actions are, or should, be taken that will affect the future of the digital images. Many of these actions will affect the longer term survival of the images and will determine if it is merely a collection of bits, something that remains fit for purpose and usable.



8.4.i Creation

The ability to approach the curation and preservation of digital images with some degree of confidence starts with the decisions taken at the point of creation or capture. It is at this point that decisions will be made on the quality of the image, the colour resolution, the metadata that is to be captured and created, the technical standards and metadata standards to be used – or not used, as the case may be. It is here where decisions about software and hardware platforms are made that may impact on how embedded the images are within a particular platform. It is also here that rights are cleared (or should be cleared) and that care must be taken to ensure that any licence agreement embody a right to preserve the image – that is, to take the necessary preservation actions to ensure future accessibility.

These decisions are likely to be driven by user and/or creator requirements, the costs involved, and the resources in terms of expertise, people and equipment that are available. Decisions may also be driven by a selection policy. It is important that all those who create and capture digital images record information on the decision making process in order that these decisions can feed into the ongoing sustainability and preservation of the resources. To this end **JISC may wish to consider if it has a role in defining benchmarks against which digital imaging projects measure their compliance with standards and best practice; and tools to capture their project processes.**

The type of information that should be recorded at this stage includes:

- Technical and metadata standards used
- Details of any pilot or feasibility testing

- The process and methodology by which images were captured or created, including equipment used, quality standards applied, and software used
- Quality assessment procedures
- Subsequent management or transformation actions such de-speckling
- Naming system used (unique IDs etc.)
- Rights and legal actions undertaken

Those responsible for preservation may or may not have control over the creation or capture of digital image collections. If they don't then the preservation repository should ascertain the standards and methods used to create the images and to assess what actions might be required to bring the images into the managed environment of the repository and, if necessary, to assess the likely costs. This assessment process will feed into decisions made at transfer and ingest, and indeed into the longer term curation of digital image collections. Depending upon the decisions made at this stage, the suitability and viability of the images for longer term access and preservation will be determined. Wherever possible, repositories should seek to liaise with content creators through the process of creation.

8.4.ii Transfer / Ingest

Once a collection has been selected for long term preservation (and not all images necessarily will depending upon the decisions made at creation, and the selection policy of the repository) then a range of actions will need to be undertaken to ensure the successful transfer to the preservation environment. These will need to be underpinned by a range of policies including the selection policy of the repository (informed by its role and responsibilities), and the preservation policy which describes the requirements for content to be ingested. These policies are likely to be driven by its designated user community, or in some cases the designated creator community – repositories policies may be driven by a remit to collect *from* a community, or by a remit to collect *for* a community. Acquiring some understanding the practices of the content creators, and of subsequent content users can only assist the repository to do a better and more efficient job.

Preservation policies should define the collecting community, the range and type of material it seeks to preserve, and its preservation responsibilities. It should state the levels of service offered and any dependencies involved, such as requiring deposit of particular formats – these might be separated into recommended and preferred. Repositories may well wish to exclude certain types of content of formats if they do not reach the accepted standards, or to limit the preservation actions that may be undertaken. The policies should also outline the metadata required to accompany the digital image collection.

The point of transfer is the first, and often only, point of substantial contact between the content owner and the preservation repository. This is a crucial opportunity to provide feedback to the creator that may improve the preservation characteristics of later submissions. It is the repository's best opportunity to collect resource discovery and administrative metadata needed to manage the images in the long-term, but more importantly, it is the repository's only real chance to establish a formal legal agreement to govern the long-term care of the images.

On transfer for preservation the following steps should be taken:

- Check that the image files are in an acceptable preservation format – if not they need to be migrated to an acceptable format if the producer has technology available to do so;

- Carry out a virus check to ascertain that the files are virus-free before transfer; if files are infected they should be replaced with uninfected versions;
- Create checksums for the individual image files; if the files are to be transferred on a portable medium (eg. CDs, DVDs) create a checksum for each instance of the medium;
- Prepare the documentation and metadata that must accompany the image files and which are necessary to ensure that they remain accessible and useable over time;
- Digital files are ingested by the preservation service initially onto a staging/processing server; this includes but is not limited to the following steps:
    - media and file readability check
    - check file counts/formats/names against any documentation provided
    - compare checksums to ensure no data has been corrupted
    - check documentation is adequate for data provided
    - copy the data to appropriate place on server
    - data validation and consistency checks
    - create receipt to send depositor

It is at this point that a collection of images might be rejected if they don't meet the necessary quality or other standards. It is important to bear in mind that not everything must or should be preserved.

8.4.iii Curation / Preservation

Once the initial set of actions have been undertaken at transfer then the process of curation proper may start. It is at this point that decisions must be taken on the level of format normalisation that occurs, the number of copies that will be created and preserved, the assignment of persistent identifiers, and verification and integrity checks that will be carried out. It is also important at this stage to define access rights for future use and the method of providing access.

It is recommended that this process is carried out on dedicated server separate from the primary preservation repository. Actions to be carried out include:

- Carry out any preservation actions that are necessary to ensure long-term usability of the image, for example any migration or normalisation actions that might be necessary;
- Validate any conversions/migrations that have been performed;
- Assign persistent identifiers
- Assign version number(s)
- Prepare dissemination version(s) of image as appropriate;
- Ensure that preservation actions are adequately documented (prepare an audit trail);
- Move preservation copies to preservation repository and check for data integrity;
- Copy original (source) bitstreams to the preservation server;
- Ensure file and directory structures are consistent;
- Carry out ongoing verification of the objects in the repository using checksum techniques.

Ideally, the preservation repository will be a stand alone system not connected to any external network. This is the best way of ensuring that data authenticity, integrity, and security is not compromised.

It is recommended that repositories responsible for long-term preservation create a set of policies and procedural documents that provide in some detail the processes they will undertake to ensure future access to the image collections. The procedures documents (what might be termed ingest manuals) will document the actions that will be undertaken for each set of images, and will specify the audit trail to be documented during the ingest/curation process. It is important that clear and transparent procedures are followed at all stages. This will ensure that it is always possible to retrace the steps taken should that prove necessary in the future.

8.4.iv Access and Use

The primary aim of preservation is to ensure continued access to digital content. The key concept here is fitness for purpose. There is little point (other than historic or to allow replication of research) in continuing to provide access to an image or set of images in a way that no longer meet the needs of the community of users. It is therefore to be recommended that repositories monitor access and use of their collections, and the behaviour of their users, especially with regard to the software they are using to manage their images for research and/or teaching and learning purposes. It may well become necessary to create new delivery versions (AIP) for users. In these cases it is recommended that in order to maintain the integrity of the delivered version of the images, repositories use the master version held in the preservation system to re-create a new version for access and delivery. This process should be properly recorded and verified as before.

Monitoring fitness for purpose in this way may also lead the repository to the decision to withdraw the image from its collection. If its quality and usefulness have deteriorated to such an extent that it no longer serves its purpose, then it may be necessary to withdraw a single image or set of images, or to make a decision to re-create them (where this is possible). **It would be useful for JISC to commission research into when re-creation might be an appropriate strategy. This could include an assessment of the costs of continuing to maintain an inadequate image, against the cost of re-creating it.**

Best practice would include the following practices:
Monitor use of images
Monitor user practices and use of software
Creation of new version of delivery versions (with appropriate integrity checks)
Re-create image if appropriate (suitable only for digital representations of an analogue image)
Assign version numbers as necessary
Fully document this process

8.4.v Technological Obsolescence

Technological obsolescence and media degradation are the central problems to overcome when planning for the long-term preservation of digital image collections. A process of monitoring the collections and the software and hardware in which they are contained for possible obsolescence and degradation is necessary, alongside a set of procedures which explain the necessary actions needed to overcome these. When action is deemed necessary then a programme of media refreshment, further migration or invoking an emulation process should be undertaken and fully documented. The same process of validation and integrity checking should be undertaken at all stages in this process.

Underpinning this must be a technology watch process to manage the risk as technology evolves and to keep up to date with new technologies such as new emulators which may be emerging. A process of risk assessment of the image content is necessary – understanding what content is held, how it is held, and what the risks are to it due to format or media obsolescence is essential. A Risk Assessment Policy such as that recommended by the ERPANET Risk Communication Tool would be a useful aid to this process.

8.4.vi Reject or Withdraw Images

The life-cycle model highlights three places where this process may happen, and these have been further explained in the text above. It is vital that any preservation strategy builds in the ability to reject or withdraw content at designated places in the life-cycle and that these are transparent and fully documented. Rejecting or withdrawing a collection is not a decision to be made lightly (unless there are concrete legal reasons for doing so of course) and should therefore be underpinned by well-developed policies and procedures.

When images are withdrawn they may be physically deleted from the repository, but the better practice already followed by many repositories, is to maintain the original images, but to mark them as no longer available or as superseded by a newer version. A reference or link to the new version is recommended.

8.5 Organisational Models

While galleries, libraries and museums undertake established roles in the preservation of analogue images, the allocation of responsibilities for the preservation of digital image material is still evolving. Image collections are collected, stored, and delivered within a variety of organisational settings, some of which are better equipped than others to meet the functional and non-functional requirements for the long-term preservation of digital images.

Within the preservation community there is a growing awareness that in the digital world responsibility for preserving information will need to be distributed in new ways. The way forward envisioned by many is to disaggregate the tasks undertaken by a digital repository, so that not all repositories need undertake all tasks.

> Fundamental to implementing this disaggregated model is the logical separation of the content and service components…. This separation allows for distributed open access content repositories to be maintained independently of value-added services fulfilled discretely by multiple service providers. Crow (2002a)

Digital preservation could be seen as one of these 'value-added' services, and could be provided in a number of ways, as suggested in the JISC Continuing Access and Digital Preservation Strategy 2002-5 (Beagrie, 2002, p. A13). Preservation of images could take place within an institutional repository, or be part of a collaborative service with a group of other repositories, or could be undertaken by an external agency or a national service of some kind.

In the JISC e-infrastructure, a number of organisational models for the provision of archival e-print repositories could develop. These models are not mutually exclusive, and disaggregated provision of archival repository functions does not necessarily require the establishment of national services. Institutions, or consortiums of institutions could provide their own preservation services, while commercial solutions could also play a role.

8.5.i Single Institution Image Repository

Image repositories operated within larger institutions may be in a position to undertake the full range of activities to meet OAIS functional requirements and the attributes of a TDR. Systems such as DSpace and Fedora offer institutions an off-the-shelf open source solution that could be implemented and used. However, the skills required to meet the TDR requirements are significant and this route requires commitment of both people and money if it to be successful. This is particularly true if a repository takes on responsibility for the preservation of vector graphics.

8.5.ii Image Repository with Specialist Support

Otherwise self-contained image repositories may need, or prefer, to call upon external services with specialist expertise in digital preservation. In the JISC e-infrastructure JISC Services such as the CCLRC, AHDS, or ESDS, further supported by the DCC, could provide these services.

8.5.iii Image Repository with Outsourced Preservation Services

Following the model currently under development by the SHERPA DP project (**http://www.ahds.ac.uk/about/projects/sherpa-dp/index.htm**) preservation planning and activity could be outsourced to an external organisation which then works in partnership with the image repository or a group of image repositories to provide an overall OAIS compliant and TDR compliant service.

8.5.iv Outsourced Image Repository Services

An individual academic, project, interest group or institution could make use of an external repository service. More than one supplier of image repository services may emerge, such as those on offer for the e-prints. Outsourcing image repository services could prove a cost-effective solution for the image collections of smaller institutions, projects and individual academic staff, and for the more complex image collections such as exists for vector images.

8.6 Disaggregated Models

Given the complexity of preserving digital image collections, and the scarcity of a pool of skilled staff trained in digital image preservation it is recommended that JISC consider the development of a disaggregated but networked model for the long-term preservation of digital images of all kinds.

The disaggregated model recommended here suggests separating out issues to do with the content, particularly liaison with the creators of digital images, the collection and retention policy, and assessment of image collections submissions, from the technical management and delivery of the image – the infrastructure services.

The former is likely to require subject expertise and experts in order to fully appreciate and understand the requirements of both start and end users of digital images, and to be able to assess the quality and form of the content. Subject expertise is also required to understand metadata and associated documentation requirements and, most crucially, to assess fitness for purpose of the images. These experts should understand the technical requirements for managing and providing access to image collections, and be able to convey these requirements to systems developers and engineers, but they need not have practical experience themselves in developing technical systems or of data management and preservation.

The Infrastructure Services, namely the long-term data management, archival storage and access functions that are required to operate the repository, and the systems development and management that supports these activities could be divided up in various ways. Most obviously, there are many commercial and non-commercial organisations capable of providing the preservation functions. In an institutional setting, it may be that computer services will take on responsibility for archival storage, but that ingest, data management and access might be controlled by library services. Alternatively, it may be that groups of subject specialists take on responsibility for content matters, working with a repository service provider who provides the technical infrastructure services that support access and delivery and long-term preservation.

At one extreme all the constituent parts may be located in the same organisation, although perhaps spread across a number of sections of that organisation. At the other extreme, they may be distributed across multiple organisations. The need to ensure that work practices are compatible, communications and management are efficient, and services are technically interoperable will place some practical limits on the disaggregation of an image repository, but there is still considerable scope for a variety of solutions to emerge**.**

Infrastructure and specialist subject support services may be provided by a single organisation in some situations. The OCLC Digital Archive (**http://www.oclc.org/digitalpreservation/services/archiving/digital/**) offers this type of unified service, while within the JISC e-infrastructure existing services such as the AHDS, the Economic and Social Data Service (ESDS), EDINA or MIMAS could provide a similar combined service.

JISC is recommended to investigate the feasibility and form of disaggregated models for the preservation of digital images.

9. Assessing Preservation Costs

Brian Lavoie wrote in 2003 "Digital preservation can hardly be classified as a new topic anymore, yet we still find ourselves not very far from the beginning in terms of exploring its economic ramifications. No systematic study of the economics of digital preservation has yet emerged."[121] In late 2005 little has changed.

Concerns about the cost are an important consideration when making decisions about preserving digital resources, and are often seen as a significant barrier. However, there is little concrete evidence that can be used as a guide and there is a general assumption that the costs will be much higher than the costs of preserving analogue materials. Because the few existing operational digital preservation systems are too recent to be considered mature there is little concrete data available about costs of both establishment and operations. For this reason it is hard to justify the assumption about costs – it seems to be felt instinctively but cannot yet be proved or disproved.

A search on 'costs' at the National Library of Australia's PADI database reveals some 90 articles with 'costs' as a keyword, in the period 1998-2005. Very little of this work is of a practical nature, the bulk of the articles being conceptual and general discussions of the issue, or passing reference in papers on other topics. It is difficult to extract either useful costing data or formulae from these papers.

Some institutions with some practical experience have tried to quantify the costs of their digital preservation programs generally. While these costings may be accurate representations of the costs for the institutions which have carried out the exercise, it is hard to see their applicability outside those institutions. Cornell University has worked out the costs of running their arXiv service as US$0.94 per digital item per year. The UK National Archives has carried out a similar exercise for its own preservation service and came up with a figure for ingest costs of GBP18.76 per file per year. The staggering differences in these two figures should warn us to be very careful about drawing any conclusions for other institutions.

In a more theoretical mode, the British Library has investigated costs in the context of its 'lifecycle' model for resources. Their formula for the costs associated with acquiring and preserving a monograph is:

$$K(t) = s + a + c + pl = hl + p(t) + cs(t) = r(t)$$

where

s = selection cost;
a = acquisition cost (excluding purchase price);
c = cataloguing cost;
pl = initial preservation cost;
hl = handling costs;
p(t) = the likely preservation cost over time;
cs(t) = the collection storage cost over time;
r(t) = the likely retrieval and replacement cost over time.

Although useful as a way of conceptualising the various cost components, the problem with this formula is that the one of the specific costs we are most interested in, preservation costs over time, is an estimate. The formula may be of value to the

[121] Brian F. Lavoie, *The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making*, OCLC Online Computer Library Center, Inc., April 2003, 42.

BL for estimating total costs associated with its acquisition and management of a resource over time, but would seem to be of little use for calculating actual digital preservation costs. The BL itself warns that it is too early and there are too many unknowns to make accurate long-term costings.

Another conceptualisation has been done by the Dutch National Library (KB) on the basis of its experience with implementing a Universal Virtual Computer approach to preserving PDF files.[122] The formulae developed by the KB are used to compare migration and emulation costs. They are:

1. Migration: K(t,a) = h(t,a) + m(t,a)

where

K(t,a) = total cost of holding *a* objects for a period of *t* years,;
h = storage costs;
m = migration costs.

2. Emulation: K(t,a) = h(t,a) + E + e(t)

where

K(t,a) = total cost of holding *a* objects for a period of *t* years;
h = storage costs;
E = costs of setting up the emulation virtual machine;
e(t) = costs of emulation over time.

Using these formulae the KB calculates that migration costs are significantly higher than emulation costs. However, I believe there are important problems with the KB calculations. Firstly, its view of migration is very limited in that it sees this approach as requiring successive migrations to newer data formats every 3 – 5 years (see Chapter 7); secondly, the KB has chosen emulation as its preservation approach and needs to justify this choice.

A recent paper by Emeritus Professor Laurie Hunter of Glasgow University treats digital preservation, interestingly, as an intangible, ie a form of capital which has no physical existence, unlike tangible capital in the form of plant, buildings and equipment.[123] Professor Hunter then places digital preservation inside the larger context of an institution's overall business strategy by arguing that we need to understand not just the costs but also the value of digital preservation. A report on the same workshop at which Professor Hunter gave his paper concludes that "as yet, we have very few concrete answers. As such, much more work must be done in determining useable cost models, in identifying practical benefits, and establishing the value of digital preservation."[124]

The OAIS model encourages us to embed digital preservation within the broader strategic goals of any organisation and not to undertake preservation in isolation from other business activities. It identifies the major components of a functional digital preservation system/service – ingest, preservation planning, data management,

[122] Erik Oltmans and Nanda Kol, "A Comparison Between Migration and Emulation in Terms of Costs", *RLG DigiNews*, vol. 9, April 2005.

[123] Laurie Hunter, "Digital Preservation as an Intangible Asset: An Overview", *Digital Preservation Coalition & Digital Curation Centre Workshop on Cost models for Preserving Digital Assets*, British Library, London, July 26, 2005, 2. Available from: **http://www.dcc.ac.uk/docs/Wksppaper.pdf** [last checked 17 November 2005].

[124] Maggie Jones, *Report for the DCC/DPC Workshop on Cost Models for preserving digital assets*, DPC July 2005. Available at **http://www.dpconline.org/graphics/events/050726workshop.html** [last checked 17 November 2005].

archival storage, administration & management, and access - but does not give any further guidance for costing such a service.

In assessing or measuring costs it would be useful to first identify the range of cost components that contribute to the overall cost of preserving a digital object. In the Cornell arXiv.org work (see above) the costs were broken down into cost categories which were further subdivided into cost centres. The resulting hierarchy was:

Cost categories:	startup costs
	ongoing costs
	contingency costs
Cost centres:	capital costs
	direct operating costs
	overhead

The UK National Archives modelled digital preservation costs by taking into account all the cost inputs involved in preservation. The TNA cost elements were broken down according to the OAIS model and consisted of:

Ingest:	selection and evaluation
	transfer
	pre-accessioning process
	cataloguing
	loading
Data and Storage	Management
	systems administration
	hardware costs
	software costs
	media management
Preservation intervention.	

In 2003 Shelby Sanett published an article on the costs of preserving authentic digital records.[125] Her modelling is applicable to all digital objects, not just records, and might be a useful starting point for modelling digital preservation costs generally. The table below is based on Table 4 in the article cited above:

Costs for Acquiring and Preserving Electronic Records	
Capital Costs	• Software development • Hardware (for preservation processing) • Research and development • Facilities • Interface design for processing electronic records
Direct Operating Costs	• Identify potential records • Evaluate/Examine (negotiate intellectual property issues and rights) • Acquire records (staff and purchase or royalty payment) • Establish inventory record

125 S. Sanett, "The Cost to Preserve Authentic Electronic Records in Perpetuity: Comparing Costs across Cost Models and Cost Frameworks", *RLG DigiNews*, August 2003, Volume 7, Number 4. Available from: **http://www.rlg.org/legacy/preserv/diginews/diginews7-4.html#feature2** [last checked 212 November 2005].

	• Process (prepare for preservation, confirm authenticity/integrity of record) • Produce metadata • Preserve (select and implement appropriate strategy) • Storage (container/other) • Maintenance (refresh/migrate) • Monitor • Evaluate • Delete
Indirect Operating Costs (Overhead)	• Indirect staff (supervision, clerical support, benefit times, training times, unallocated times) • Facilities (rent, utilities, off-site storage of records) • Amortization of capital costs • General and administrative (human resources, accounting, funding development and grant writing, staff training and professional development, partnerships with other institutions, policy development)

In view of the significant gaps in the literature of costing it does not seem possible to recommend any particular approach or formulae for assessing or estimating the costs of preserving digital assets. A generalised approach based on proper identification of all the cost components involved in preserving a digital objects might be the best approach until more specific and detailed cost studies appear in the literature.

JISC Digital Image Study

Appendix 1: User surveys

A1. Targeted surveys

Respondent 1 – **Historian and researcher**

1. What is your main area of interest relating to digital images?

Preservation and preparation of digital images for preservation. Also I direct a research project creating c.200,000 digital images.

2. How important are digital images to your work?

	Essential	X	Important		Not too important		Irrelevant

3. How would you describe the sort(s) of images that you use most regularly?

Poor quality

4. What purpose(s) do you use the images for?

	Lectures		Handouts	X	Research	X	Publication
	Exhibitions		Private study		Projects		VLE
	Other						

5. What are the main problems with accessing digital images?

	Don't know where to look		Not enough support
X	Lack of subject relevance		Limited access to technology
X	Poor quality of images	X	Lack of metadata
	Concerns about copyright		Wrong formats for my use
	Other		

6. Where would you like to access digital images?

X	Own PC		Library		Online (WWW)		CD/DVD
X	Repository		Intranet		College server		VLE
	Other						

7. Have you thought about the long-term issues of digital image availability and, if so, in what way does this affect you?

Yes. As a service provider who has been following a migration-based strategy for the preservation of digital data, digital images present a different problem. There are many fewer options for digital images and other strategies may have to be pursued in the future.

8. What experiences do you have of using images in a digital repository?

Very little

9. What do you think might be the main issues of images stored over long periods of time?

	Quality of images		Copyright
	Provenance		Ability to search
	Selection of images to preserve		Availability in the future
X	Changes in technology		Loss of resources
	Other	**Query: aren't availability in the future and loss of resources the same thing? Also, I would have thought that "selection of images to preserve" was a little nonsensical as all images which have a particular quality threshold should be preserved.**	

10. Can you think of other issues that should be explored?

Well, an extension to provenance. Something along the lines of construction metadata, i.e., to explain what processing, i.e., dithering, despeckling etc., has taken place from the original image to the preservation version, and what algorithms etc., have been implemented. (Basically how far is this from an "original" digital image.

11. What are the most important aspects of a digital image for your work?

	Format	X	Subject
	Provenance		Pixel quality
	'Findability' (Metadata)		Flexibility of use
	Other		

12. Who do you think should care for digital images in the long-term?

	One national body	X	National subject-based bodies
	Regional bodies		Institutions
	Institutional departments		Individuals
	Other		

13. How should access to preserved images be controlled?

	Free access for all		Registered users only

	Access to educational users only		Different levels of access
	Use by paid subscription		
	Other	**Free access to registered users**	

14. How should a preservation system control quality?

	Anything should be accepted	X	Technical requirements must be met
	Signed licence agreements		Legal checks
	Peer review in advance	X	Adherence to metadata schemas
X	Agreed quality standards		User feedback/ratings added
	Other		

15. Is there anything else you think should be considered in the process of preserving digital images?

Well, what's not mentioned here is structural metadata which may or may not allow one digital image to be related to another in a meaningful way. The (unspoken) emphasis in this survey is that digital images are in many respects stand-alone, but many digital images of manuscripts, and books and other printed materials are sequential. The implicit structure of these things is important, and is often ignored, as is the fact that books may have multiple sequences of pages (Arabic and roman numerals) and mss have odd page sequencing.

We have a collection of some 300 books covering over 150,000 pages. Each image scan is uniquely identified, and each book is given a folder. The names of the pages of each "book page" are not "sequential" in any definition of the term, thus the only way we know the sequence of pages is sorting by date of creation. Imagine the problems….

Respondent 2 – **University curriculum advisor**

1. What is your main area of interest relating to digital images?

Use of digital images to enhance online learning materials

2. How important are digital images to your work?

	Essential		Important	√	Not too important		Irrelevant

3. How would you describe the sort(s) of images that you use most regularly?

4. What purpose(s) do you use the images for?

√	Lectures	√	Handouts		Research	√	Publication
√	Exhibitions		Private study		Projects		VLE
	Other	Creating learning materials					

5. What are the main problems with accessing digital images?

√	Don't know where to look	√	Not enough support
	Lack of subject relevance	√	Limited access to technology
	Poor quality of images		Lack of metadata
√	Concerns about copyright		Wrong formats for my use
	Other	Time	

6. Where would you like to access digital images?

	Own PC		Library	√	Online (WWW)		CD/DVD
	Repository		Intranet		College server		VLE
	Other						

7. Have you thought about the long-term issues of digital image availability and, if so, in what way does this affect you?

Have considered it and one major issue with the use of Images is how to continue to use them if license conditions change. For example, if a College has a license for say the EIG and some of the images are used to create learning materials then subsequently the College doesn't renew its licence what status do the images embedded in the learning materials have?

8. What experiences do you have of using images in a digital repository?

Some experience of JORUM.

9. What do you think might be the main issues of images stored over long periods of time?

	Quality of images	√	Copyright
	Provenance	√	Ability to search
	Selection of images to preserve		Availability in the future
√	Changes in technology		Loss of resources
	Other		

10. What do you think might be the main issues of images stored over long periods of time?

11. What are the most important aspects of a digital image for your work?

	Format	√	Subject
	Provenance		Pixel quality
	'Findability' (Metadata)	√	Flexibility of use
	Other		

12. Who do you think should care for digital images in the long-term?

√	One national body		National subject-based bodies
	Regional bodies		Institutions
	Institutional departments		Individuals
	Other		

13. How should access to preserved images be controlled?

√	Free access for all	√	Registered users only
	Access to educational users only		Different levels of access
	Use by paid subscription		
	Other		

14. How should a preservation system control quality?

	Anything should be accepted		Technical requirements must be met
	Signed licence agreements		Legal checks
√	Peer review in advance		Adherence to metadata schemas
√	Agreed quality standards		User feedback/ratings added
	Other		

15. Is there anything else you think should be considered in the process of preserving digital images?

Respondent 3 **- Director of information & learning technology development**

1. What is your main area of interest relating to digital images?

Responsible for ILT Services which includes Learning Object Development and LRC – Repository and searching etc.

2. How important are digital images to your work?

X	Essential		Important		Not too important		Irrelevant

3. How would you describe the sort(s) of images that you use most regularly?

A broad range, some as pure eye candy to those that are intrinsic to the learning episodes

4. What purpose(s) do you use the images for?

	Lectures	X	Handouts		Research		Publication
	Exhibitions		Private study	X	Projects	X	VLE
	Other	**Use in Learning Objects which may be placed in a VLE**					

5. What are the main problems with accessing digital images?

	Don't know where to look		Not enough support
	Lack of subject relevance		Limited access to technology
	Poor quality of images		Lack of metadata
X	Concerns about copyright		Wrong formats for my use
	Other	**Lots of images on the web but copyright prevents their use – using collections is ok but you have to pay a lot for a few relevant images**	

6. Where would you like to access digital images?

	Own PC	X	Library	X	Online (WWW)		CD/DVD
X	Repository	X	Intranet	X	College server	X	VLE
	Other						

7. Have you thought about the long-term issues of digital image availability and, if so, in what way does this affect you?

We are coming to the conclusion that it might be worth building our own repository and stocking it with copyright cleared and self generated images that we know are safe to use.

8. What experiences do you have of using images in a digital repository?

Some have been difficult to download and the copyright information is restrictive and certainly is lost when there is copying of copying – pragmatic position

9. What do you think might be the main issues of images stored over long periods of time?

	Quality of images	X	Copyright
	Provenance		Ability to search
	Selection of images to preserve		Availability in the future
	Changes in technology		Loss of resources
	Other		

10. What do you think might be the main issues of images stored over long periods of time?

As above – isn't this the same question?

11. What are the most important aspects of a digital image for your work?

	Format	X	Subject
	Provenance		Pixel quality
	'Findability' (Metadata)	X	Flexibility of use
	Other		

12. Who do you think should care for digital images in the long-term?

X	One national body		National subject-based bodies
	Regional bodies	X	Institutions
	Institutional departments		Individuals
	Other		

13. How should access to preserved images be controlled?

X	Free access for all		Registered users only
	Access to educational users only		Different levels of access
	Use by paid subscription		
	Other		

14. How should a preservation system control quality?

X	Anything should be accepted		Technical requirements must be met
	Signed licence agreements		Legal checks
	Peer review in advance		Adherence to metadata schemas

	Agreed quality standards		User feedback/ratings added
	Other		

15. Is there anything else you think should be considered in the process of preserving digital images?

I think the fact they are digital makes their creation and adaptation easier – we should therefore not apply the same thinking process to this as we would when dealing with slides or prints – ease of use makes them more useable and so we need to deal with the notion of images as eye candy as well as for sound educational use.

Respondent 4 – **Director, advanced technology and information**

1. What is your main area of interest relating to digital images?

The long term preservation of our cultural heritage. Images provide a powerful vantage from which to view contemporary society; they are evocative in a way that textual and audio data sets can never be.

2. How important are digital images to your work?

	Essential		Important	x	Not too important		Irrelevant

Actually this is a difficult question for me to answer as I do not actually use images in my work they can not be said to be critical to it. What I am interested in is the long term accessibility of digital culture and image resources are an important aspect of that. So while personally I make little use of images they are extremely critical to my view of cultures.

3.How would you describe the sort(s) of images that you use most regularly?

Images of events and of technical devices for explanatory purposes.

4. What purpose(s) do you use the images for?

X	Lectures	X	Handouts	X	Research	X	Publication
	Exhibitions		Private study	x	Projects	X	VLE
	Other						

5. What are the main problems with accessing digital images?

X	Don't know where to look		Not enough support
X	Lack of subject relevance		Limited access to technology
	Poor quality of images	X	Lack of metadata
X	Concerns about copyright		Wrong formats for my use
	Other		

6. Where would you like to access digital images?

X	Own PC		Library	X	Online (WWW)		CD/DVD
X	Repository	X	Intranet		College server	X	VLE
	Other						

7. Have you thought about the long-term issues of digital image availability and, if so, in what way does this affect you?

The key issues from my point of view are accumulation (e.g. collection of images), appraisal, provenance, annotation, but above all else IPR. The accessibility of images will impact on the ways that my colleagues can teach and the kinds of learning that my students can undertake.

8. What experiences do you have of using images in a digital repository?

Little or none. I have been responsible for the design of systems for managing images.

9. What do you think might be the main issues of images stored over long periods of time?

4	Quality of images	2	Copyright
1	Provenance	6	Ability to search
9	Selection of images to preserve	3	Availability in the future
5	Changes in technology	8	Loss of resources
	Other		

10. What do you think might be the main issues of images stored over long periods of time?

For raster image formats the problems of long term accessibility from a technical vantage are relatively simple in comparison to those associated with other data types. From my point of view the greatest challenge that we face is provenance and IPR.

11. What are the most important aspects of a digital image for your work?

6	Format	2	Subject
1	Provenance	5	Pixel quality
4	'Findability' (Metadata)	3	Flexibility of use
	Other		

12. Who do you think should care for digital images in the long-term?

1	One national body		National subject-based bodies
	Regional bodies		Institutions
	Institutional departments		Individuals
	Other		

13. How should access to preserved images be controlled?

X	Free access for all		Registered users only
	Access to educational users only		Different levels of access
	Use by paid subscription		

	Other	

14. How should a preservation system control quality?

	Anything should be accepted	X	Technical requirements must be met
	Signed licence agreements	X	Legal checks
	Peer review in advance	X	Adherence to metadata schemas
X	Agreed quality standards	X	User feedback/ratings added
	Other		

15. Is there anything else you think should be considered in the process of preserving digital images?

Annotation. It is essential that users of images be able to annotate them in different ways and for these annotations to be propagated across collections and space/time.

Respondent 5 – L**ibrary and information services specialist**

1. What is your main area of interest relating to digital images?

As a librarian to be able to provide these for students and staff to use – particularly in PowerPoint.

2. How important are digital images to your work?

	Essential		**Important**		Not too important		Irrelevant

3. How would you describe the sort(s) of images that you use most regularly?

Staff and students use slides of works by individual artists and designers.

4. What purpose(s) do you use the images for?

	Lectures		Handouts		Research		Publication
	Exhibitions		Private study		Projects		VLE
	Other	Students use for presentations + staff for lectures					

5. What are the main problems with accessing digital images?

	Don't know where to look		Not enough support
	Lack of subject relevance		Limited access to technology
	Poor quality of images		Lack of metadata
	Concerns about copyright		Wrong formats for my use
	Other		

6. Where would you like to access digital images?

	Own PC		Library		Online (WWW)		CD/DVD
	Repository		**Intranet**		**College server**		VLE
	Other						

7. Have you thought about the long-term issues of digital image availability and, if so, in what way does this affect you?

No

8. What experiences do you have of using images in a digital repository?

I've only used digital images from websites like AHDS or Scran.

9. What do you think might be the main issues of images stored over long periods of time?

	Quality of images		Copyright
	Provenance		Ability to search
	Selection of images to preserve		Availability in the future
	Changes in technology		Loss of resources
	Other		

10. What do you think might be the main issues of images stored over long periods of time?

11. What are the most important aspects of a digital image for your work?

	Format		**Subject**
	Provenance		Pixel quality
	'Findability' (Metadata)		Flexibility of use
	Other		

12. Who do you think should care for digital images in the long-term?

	One national body		**National subject-based bodies**
	Regional bodies		Institutions
	Institutional departments		Individuals
	Other		

13. How should access to preserved images be controlled?

	Free access for all		Registered users only
	Access to educational users only		**Different levels of access**
	Use by paid subscription		
	Other		

14. How should a preservation system control quality?

	Anything should be accepted		**Technical requirements must be met**
	Signed licence agreements		Legal checks
	Peer review in advance		Adherence to metadata schemas
	Agreed quality standards		User feedback/ratings added
	Other		

15. Is there anything else you think should be considered in the process of preserving digital images?

Respondent 6 – **Artist, lecturer and researcher**

1. What is your main area of interest relating to digital images?

As an artist, Lecturer, Researcher and Director of DACS, Design and Artist Copyright Society.

2. How important are digital images to your work?

√	Essential		Important		Not too important		Irrelevant

3. How would you describe the sort(s) of images that you use most regularly?

Images created by me or the artists I select from the digital shows I curate.

4. What purpose(s) do you use the images for?

√	Lectures		Handouts	√	Research	√	Publication
√	Exhibitions	√	Private study	√	Projects		VLE
	Other						

5. What are the main problems with accessing digital images?

	Don't know where to look		Not enough support
	Lack of subject relevance		Limited access to technology
	Poor quality of images		Lack of metadata
√	Concerns about copyright		Wrong formats for my use
	Other		

6. Where would you like to access digital images?

	Own PC		Library	√	Online (WWW)		CD/DVD
	Repository		Intranet		College server		VLE
	Other						

7. Have you thought about the long-term issues of digital image availability and, if so, in what way does this affect you?

That the resolution and the file format remains applicable!

8. What experiences do you have of using images in a digital repository?

Through JIDI and AHDS Visual Arts

9. What do you think might be the main issues of images stored over long periods of time?

√	Quality of images	√	Copyright
	Provenance		Ability to search
	Selection of images to preserve		Availability in the future
	Changes in technology		Loss of resources
	Other		

10. What do you think might be the main issues of images stored over long periods of time?

11. What are the most important aspects of a digital image for your work?

√	Format		Subject
	Provenance	√	Pixel quality
	'Findability' (Metadata)	√	Flexibility of use
	Other		

12. Who do you think should care for digital images in the long-term?

√	One national body	√	National subject-based bodies
	Regional bodies		Institutions
	Institutional departments		Individuals
	Other		

13. How should access to preserved images be controlled?

	Free access for all		Registered users only
√	Access to educational users only	√	Different levels of access
	Use by paid subscription		
	Other		

14. How should a preservation system control quality?

	Anything should be accepted		Technical requirements must be met
√	Signed licence agreements	√	Legal checks
	Peer review in advance	√	Adherence to metadata schemas
√	Agreed quality standards		User feedback/ratings added
	Other		

15. Is there anything else you think should be considered in the process of preserving digital images?

Respondent 7 – **Director, commercial online directory**

1. What is your main area of interest relating to digital images?

The conversion, presentation and preservation of images in the context of a national information resource.

2. How important are digital images to your work?

✓	Essential		Important		Not too important		Irrelevant

3. How would you describe the sort(s) of images that you use most regularly?

The images are scanned, photographed or digitally created representations of contemporary artists' work.

4. What purpose(s) do you use the images for?

✓	Lectures	✓	Handouts	✓	Research	✓	Publication
	Exhibitions		Private study	✓	Projects		VLE
✓	Other	**we use the images to promote our own services.**					

5. What are the main problems with accessing digital images?

	Don't know where to look		Not enough support
	Lack of subject relevance		Limited access to technology
	Poor quality of images		Lack of metadata
	Concerns about copyright		Wrong formats for my use
✓	Other	**no problems, as we are the image provider ourselves**	

6. Where would you like to access digital images?

	Own PC		Library	✓	Online (WWW)		CD/DVD
	Repository		Intranet		College server		VLE
	Other						

7. Have you thought about the long-term issues of digital image availability and, if so, in what way does this affect you?

Backup and long-term storage is a logistical issue which is manageable. The viewable size and resolution of display devices is still increasing steadily which may become a problem if older images do not have a sufficiently high resolution. They would have to be recreated (e.g. re-scanned) or digitally enlarged to be available in a higher resolution. This in turn increases the risk of high-resolution images being used for purposes not intended. Any type of

watermarking or download/copy protection is useless in our experience, clear copyright statements are essential.

8. What experiences do you have of using images in a digital repository?

Good ones, we are a digital image repository.

9. What do you think might be the main issues of images stored over long periods of time?

✓	Quality of images	✓	Copyright
✓	Provenance	✓	Ability to search
	Selection of images to preserve		Availability in the future
	Changes in technology		Loss of resources
	Other		

10. What do you think might be the main issues of images stored over long periods of time?

Resolution too low for effective future use (see above)

11. What are the most important aspects of a digital image for your work?

	Format		Subject
	Provenance	✓	Pixel quality
✓	'Findability' (Metadata)		Flexibility of use
	Other		

12. Who do you think should care for digital images in the long-term?

	One national body		National subject-based bodies
	Regional bodies		Institutions
	Institutional departments		Individuals
✓	Other	**depends on the type of images and their use**	

13. How should access to preserved images be controlled?

	Free access for all		Registered users only
	Access to educational users only		Different levels of access
	Use by paid subscription		
✓	Other	**depends on the type of images and their use**	

14. How should a preservation system control quality?

	Anything should be accepted		Technical requirements must be

			met
	Signed licence agreements		Legal checks
	Peer review in advance		Adherence to metadata schemas
	Agreed quality standards		User feedback/ratings added
✓	Other	**depends on the type of images and their use**	

15. Is there anything else you think should be considered in the process of preserving digital images?

A2. Broad survey

This section presents the results of 101 online survey responses taken between Tues 25th Oct and Monday 14th November.

Responses alphabetically by main area of respondent interest:

Aerial imagery
Storage media and retrieval means, i.e. via metadata; Offsite storage, back up media format, security of decommissioned storage devices.

Aerial photography
Vertical aerial photography captured digitally and on film then scanned.
Quality of images is important. Most important is longevity of the images (no corruption of data and retrievable)

Aerial photography and maps
In order of importance (I think)- access (software reliability, ease of use and longevity; copyright; file size and storage; metadata; imaging, traditional AND digital QUALITY - without a doubt. I am so disheartened when looking at sites that have clearly altered the original image so much that important information is lost. That loss may not be clear to the average user but it is very clear to imaging specialists schooled in traditional photography and THAT is primarily what is being captured and preserved at this point. The choices then become rescanning or leaving the images in their grossly altered condition, which, of course, leaves users to believe the photographer's vision has been respectfully and faithfully preserved. It's bad business. Our concern should be to introduce as little non-native artifact as possible (considering that all reformatting introduces some level of alteration). And while the actual "preservation" (keeping technology from destroying file data) is ultimately at the top of the list of preservation concerns, those files are perfectly worthless if captured and altered improperly. Followed closely by access (that's the point, isn't it?). Metadata seems to be a hot topic for many lately. While I do not dispute the value, I do question the extent that some are willing to go to preserve information that may or may not be of any real value a year from now. I am not convinced that metadata is the do all and end all of long-term preservation.

Africa and Asia
Copyright, Quality and Access

Archaeology
Metadata including copyright info is clearly as important to preserve as the image itself. Most important is probably ensuring that the file formats and media used for archiving remain current. It's important that the files be preserved using a non-proprietary lossless format.

Archaeology
Preservation of digital images is not different from any other digital source. The media to store data are the major problem. DVD's and CD's are about the worst thinkable. WORM-disks are said to be relatively reliable, but there are too many 'standards'. The promotion activities of safe storage these are not worth mentioning. Storing on hard disks is common practice to day, is not a solution, UNLESS backups are stored on remote servers as well.

Archaeology: Classical archaeology.

I assume you are talking only about images born digital. If not, I would have a different set of comments. Metadata is probably most important because, absent good metadata (resource discovery metadata, not documentation metadata), the images cannot be found. Copyright issues are obviously important, but I suspect that copyright for the images made available in digital form can be made non-restrictive if action is taken quickly. Generally speaking, the creators should still be alive. Quality is critical, but it depends on the original owner more than the subsequent archives. Access must be over the web. I think the images should, as much as possible, be archived in the original format, not JPG or other compressed formats. Web serving needs to be carefully "constructed" so that access - whether to small images or full-resolution ones - is well-considered for real users.

Archaeology and History
Preserving image quality, future access and software compatibility, copyright control

Archaeology, ICT, Digital Preservation
Two main issues: at the core there must be open standard / non-proprietary image formats to ensure that the image / data itself can be used and preserved over the long-term. Based around this should be concise and effective metadata standards to ensure images and collections can be discovered and accessed while also documenting aspects of the image such as copyright and migration processes etc. Use of image compression (none, lossless and lossy) in different circumstances.

Archaeology (including excavations and recording of materials)
Copyright seems to be the main area of concern. Next would be appropriate file formats allowing for future migration. Then quality, access and metadata. Raising awareness and providing support channels about preservation issues within higher education is paramount if sufficient provision is to be given to projects and Departments generating digital images in the future. This will become increasing necessary for funding applications in the near future.

Archaeology: Late antiquity
Digitising of existing slide collections, especially those created by field archaeologists and which presumably are under threat of extinction. Clarity of how a digital image might be manipulated and reused -- Creative Commons licenses for modification and re-distribution probably even more relevant for images than for text. Quality control of digital images -- too many sites being created which do not attempt to distinguish between scholarly images and 'someone with a camera' images. Peer-review applies as much to images as it does to text (especially within the field of archaeology). A level of discrimination is required and to develop this will require knowledge of provenance.

Archaeological photography in Sino American Field School of Archaeology (SAFSA) program in China
(1) copyright if it was part of a printed report,(2) Quality (3) FREE accessibility through WEB-site for research and teaching purposes. Also (1) It should be considered to arrange by typology of "CORPUS" of each similar data.

Archives
Keeping pace with the operating hardware and software; using appropriate and meaningful metadata; loss of information and image quality through migration; Long term preservation of the original format

Archivist & Records Manager

I look after our digital images and have some input into capturing images which record the activities of our organisation for future use.
Metadata collection and copyright and the suitability of the format for long term preservation.

Arts & Humanities: Images for art and humanities and social science.
Consistent quality of files and consistent metadata. Creating excellent metadata that can be shared, and used for authority control among many databases. Every institution seems to be starting from ground zero on this count.

Art & Design
quality, access, metadata

Art & Design: Architecture
The preservation and archiving of images of architectural interest taken by staff and students of my university department and the creation of an archive of images of ongoing student work
The quality and resolution of a digital image should be high enough to allow detailed study of a building/object/model etc when the original no longer exists. This is particularly the case with student work as much physical work is destroyed due to lack of space to keep models and large drawings as well as the precarious nature of students' own digital storage methods. Comprehensive metadata is consequently of immense importance, especially as an image of a student's work needs to be closely tied to a description of the student's overall project in order to make sense of the concept behind it. Ease of archiving would be a key factor for us. When we have requested students' images of their own work in the past, despite specifying required file types etc, we have been overwhelmed by a bewildering amount and variety of image file types, resolution, and sizes and an equally varying quality of accompanying information which makes archiving extremely difficult, time consuming and inaccurate. Thus the ideal would be a means for students to upload their own images and metadata which would be easy to use and accessible but which would restrict the number, resolution, file type of the images to a necessary standard and make it easy for the student to input suitable metadata of a consistent detail and standard.

Art & Design: Contemporary Art, Design and Crafts
It is difficult to separate the issues in terms of a 'risk assessment'. Obviously changes in technology could invalidate access to an image collection unless we carefully adapt standards carefully as an academic community. The advice we receive from agencies such as TASI has been hugely important in this respect would think that choosing durable formats with transferability potential for images that ensure high quality over a period is probably the most important factor. However copyright is also a complex issue in that as institutions we need to ensure that as far as possible we own the images we buy for as long as possible. If and when we manage to create a licence that will allow HEI's to have access to a copyright cleared digital image bank of contemporary art and design images either as a national resource or regional or local then we will need to ensure that the copyright landscape is durable. This second point is somewhat academic but as an optimist I await developments in this regard.

Art & Design: fashion
quality

Art & Design: printed images, woodcut and intaglio in Western European books, flugblatter, depicting the Ottoman East 1470-1700

Quality; metadata; copyright is significant, but in most cases images are contained in libraries and archives and contractual usage rights are more significant; I am also concerned with the issues directly affecting preservation. Should we work from microfilm and create our digitised access version from that second stage but durable surrogate. Given the inherent lack of permanence in most digital media, and the need for 'maintenance migration' to ensure access, a dual approach may well be the best solution. ; I am also concerned that for fine line images the standard practice of most archives - 600 dpi scanning- may be inadequate.
Other: a simple and effective descriptive system to identify elements in the image. Not ICONCLASS ! We are working on it. Congratulations, by the way, on putting out this questionnaire. I hope that something may develop from it.

Art & Design
We started to digitally photograph the summer degree show 3 years ago and we have a slide record of degree show work for the past 30 years or so. I want to ensure the digital images are still accessible to future users in the next 30 years.
Metadata, access, new technology, copyright; Costs; selection of material to preserve or all to be preserved?; what format will images be preserved in and for how long

Art History
Quality (depth of the image), metadata (without sufficient metadata for retrieval, the images might as well not exist), open access for educational and research purposes.
Other: Changing platforms for storage and access.

Art History
Access and quality. Members of academic staff should have access to the images which have been used in the courses produced by their unit in the past for future reference and not have to go through some intermediary to get them.

Art History: Ancient art
Access; Metadata; Quality

Art History: British architectural history
Archiving, Access, quality and long-term preservation

Art History: Early Christian art, reception on classical past, museum etc exhibitions.
Copyright, quality (i.e. need to be publishable), access

Art History: Eighteenth-century European painting and sculpture
Quality is essential, especially in colour values. Also clear instructions re their potential use and copyright issues. Easy search processes. Perhaps group thematically as well as by artist/period? Timely publicity to those who may wish to access the images.

Art History: Italian art 1200-1600
Quality; Access; Copyright - this needs to be sorted out as a priority as it is extremely difficult for academics to continue to foot the bill for copyright which publishers are currently refusing to do); Deterioration of quality over time

Bookbinding: History of bookbinding and conservation
As the whole point of preservation is the preservation of access, continued access is clearly the most important issue, second to which is the preservation of the quality of the original image, as loss of quality diminishes access. Preserving the attendant

metadata is essential to knowing what the image is into the future, so an image without metadata is in many cases next to useless. All these are issues which should have been resolved before preservation becomes an issue. Same with copyright really. Anyone saving images without thinking of these matters is heading for trouble. The central issue is the PRESERVATION of the digital data; We must do what we can to make curatorial care in the future as straightforward as possible otherwise it will not be taken. We also need to use discrimination in ensuring that the digital records earmarked for long-term preservation are actually worth preserving and cannot be effectively preserved in another more durable format (i.e. paper), because it is likely to be expensive.

Classics
Quality, metadata and access - I would like copyright to go away. But we need to ensure that anything digitised is stored in such a way that as far as possible we need not fear a requirement to repeat the process of converting between incompatible systems - as with slides to digitised images

History of crime and forensic medicine
Easy access to information about location of images, who holds the copyright, and cost implications for various uses. This ties in with having someone in charge of answering enquiries about such issues in a timely fashion. Costs to potential users should be kept to a minimum, and format in which images are stored updated regularly (but access to older version maintained). Images in small collections or that are little known should be digitised where ever possible.

DAM: CMS/ Digital repositories for historical materials
Metadata is key - in terms of identifying the metadata required for all future potential uses of the content. It is not economically feasible to go back and supplement metadata once captured. If the metadata is captured all the other issues should be resolved (I am assuming, of course, that metadata includes copyright and usage information); Methods of metadata capture by non experts

DAM: creating digital archives for WWW
Image resolution - preservation quality and web quality, appropriate CMS for distribution via WWW, W3C compatibility, quality metadata; Long term Sustainability of digital archives

DAM: Digital Asset Management across a large Museum - all aspects.
Metadata, media, format and quality. If the metadata is wrong, you can't resolve copyright easily. Media is crucial; we have to wean people off of CD ROM, and DVD storage. Format is important: storage as PSD files has benefits but also huge drawbacks: also, compression algorithms are not understood by the general user. Both of these impact on quality, as does lighting, composition and resolution. Some of the above note includes an extending of the issues I would raise. However I would also add:- Platform management. Growth of images naturally leads to questions of capacity planning, storage virtualisation, disaster planning and remote hosting. At the very high-end, conversion leads to image inaccuracies. (See e.g. some of the work a while ago for the National Gallery).

DAM
Evaluating and installing a digital repository at a University - need to consider preservation.
Access (if you can't access it everything else is kind of irrelevant); metadata; copyright; What application created the digital data (including version of application); operating system used; current storage medium

DAM
Manager of museum collections management system, which includes images, and makes them available over the web.
Media (eg busted CDs or DVDs) are more likely to be a problem than file formats. Keep everything live on backed-up server hard drives, and use open source (or openish) formats (such as jpg, gif, pdf) in the file system rather than anything too proprietary. I'm suspicious of storing things as binary objects inside databases- that introduces another layer of potential incompatibility. We almost exclusively use jpegs for our object images. Metadata are an issue, but not a big one for our particular object images: files can be made self-documenting by using object numbers as file names (eg GLAHM 40123.jpg). Further object data is in the accompanying databases. In the long run, open formats are hugely important: self-expiring, or other DRM-controlled formats will be a catastrophe. These must be resisted.
Other: Seeing the issue as being mainly about theoretical metadata, obscures things for the vast majority of those generating image data, who use standard consumer tools, and/or freeware such as Irfanview. Many people who do produce digital images in scientific, and museum setups are operating in under-funded setups without much IT, or documentation support. Simple kinds of good practice which help them name, organise, and appropriately store and backup images are essential. Simple and effective user-tools which help productivity, and hide complexity, are essential if you want people to adopt rigorous XML-based metadata standards.

Digital Data Development
Structural metadata; Resource discovery metadata; Longevity of digital resource; Emulation vs. Migration Preservation strategies; Need to make many people aware that preservation and back up are not the same thing, and that long-term preservation means (to all intents and purposes) permanent preservation.

Digital imaging
Rich metadata for fast indexing and high quality masters for easy generation of accessible copies; Scalable storage infra-structure.

Digital Imaging Consultant and Trainer in Heritage Sector
Responsibility and Budget. Most aspects of preservation are simple enough. If the resource has enough 'quality' now to be useful then there is no reason why this resource should not be useful into the future and is worth preserving. The only two really important factors are simply curation. "Who is responsible for the ongoing preservation"; "what budget do they have for this work" As long as there is someone who is responsible for the preservation and as long as they have a budget for the work.....it is all pretty easy really. Anything else you think should be considered: No not really..... just 'responsibility' and 'budget' with those in place, everything else can be pretty easily worked out as necessary.

Digital video and sound
Copyright and exclusivity issues. Economics of the art market.

Digitisation consultant
Media obsolescence, search engine development, quality; repurposing

Digitisation, humanities computing
Image formats and being able to access them in the future! As a whole, I think we need to engage more with the computer science behind the digitisation process. There is a lack of understanding of the way images work, and the problems with colour reproduction, file management, and version control (some comp scientists I

work with, for example, cannot believe that we recommend Tiff as a standard for storing images as the specification has changed so many times and files are subject to corruption!) I think our own misunderstanding of what digital images are and how they function may be our downfall!

Digitisation: slide copying, copying of official and personal archives, antiquarian papers and modern GPS digital surveys
Copyright for the small man and the dominance of the national institutions. Quality is certainly an issue. Website images can be a problem because their presence is often transitory. Who archives websites?

Digitisation
Work for a service that provides support for those digitising or using digital images. Issues are not just technical, but also related to usage and management: If a resource is well used, it is more likely to survive; unless someone takes responsibility for it, it is unlikely to survive. The stability of the storage media is an issue - images need to be held in multiple copies, multiple locations and migrated periodically. The longevity of the file format is also an issue - images should ideally be kept in multiple formats and migrated to newer formats periodically. Particular formats pose a threat to preservation (e.g. JPEG quality is compromised each time resaved), and use of compression is also risky. Metadata is a key issue - ensuring that the image and all the processes applied to it can be understood. Also for locating and retrieving the image - if an image cannot be found it is essentially "lost"!

Drama
Access and copyright - though maybe that's more from the perspective of someone digitising? How choices are made about what to digitise...

E-learning: course material images in all subject areas
We are responsible for maintaining an archive of all materials produced for a University.
Quality and metadata. Quality to ensure the images are fit for purpose and so if images are migrated through formats there is as little quality loss as possible. Metadata to aid retrieval but also to give you the mechanism by which adequate preservation management can take place, review, checksum, migration strategy, etc...The main barrier is copyright. Selection criteria, long term strategy for ensuring the maintenance of the images, interoperability - ensuring both the loss of metadata in any system migration an appropriate content packaging so individual images or sets of images can be imported into other types of systems (VLES, e-portfolios, external repositories, etc...)

E-learning: Learning technologies, digital repositories, digital images in teaching and learning, stereoscopic applications
Digital Rights Management, Proprietary RAW file formats; Proprietary encryption of image data within RAW file formats

E-learning: WebCT, Questionmark.
The use of images in these and other online learning environments.
Copyright very important. Usually overlooked by practitioners. Using the right software. Types of media used for storage. How to retrieve images i.e. searchable databases, metadata etc. Controlling access to archives. Training in the re-use of images. Some way of indexing the suitability of images for use in an educational environment. Could be highly controversial/ akin to censorship?

Film: archive film

Most important issues: metadata, file formats, access, quality

Film, film scripts, drama and story telling
Most important issues: Protection of copyright. Access and publicity will to some extent prevent plagiarism. Proper documentation of the image in its original context.

Film studies
Most of the issues apply that also surround preservation of images in older formats, but the need to address them perhaps seems more urgent. Providing metadata to agreed standards is particularly important, since that is the key to access.

General
Access and longevity. I would like to remind people of the availability of microfilming as a means of providing long-term storage of images and documents. It is extremely cost-effective, easily distributed and it can also be scanned to provide digital resources.

Heritage and culture archives and learning material
Copyright, access

Heritage Recording and Archiving
Metadata and then organization of the images and supporting data. A paper was presented at the recent CIPA Heritage Documentation Symposium in Torino Italy that directly addressed this issue. http://nickerson.icomos.org/steve/papers/174-Torino-ASCix.pdf

History of Art
A collection of photographs for the study of and research into the history of architecture, sculpture and medieval painting.
My main concern is that new technology might make it impossible to access digital images. I also want to stress the importance of good cataloguing, and good metadata, in providing access to the images.

History of art and design
I believe it is absolutely essential to establish a general quality standard for archival preservation of images that is adhered to by all the major institutions. In terms of metadata, I think the current standards such as Dublin Core and the VRA core are easy to implement and versatile enough, but it is still important to ensure that standards are enforced. I also think copyright has become a major stumbling block for digital image usage and preservation, that needs to be dealt with ASAP. Anything else you think should be considered: My major worry at the moment is ensuring easy migration of formats, so that the digital images of today don't become the digital limbo of tomorrow.

History of art and the electronic environment
There will always be preservation issues but access is key. Art historians and not only slide librarians should be encouraged to address the preservation process at the very least for the preservation of the illustrated lecture.

History of Art: English medieval art
Most important issues: quality, metadata (including the fullest possible documentation of the images), access and copyright in that order. Links from copyright libraries worldwide

History of Art, History of Architecture, 19th c., 20th c., Universal and International Expositions, World's Fairs, Nationalism
Copyright and access because they are both legal issues which would need common regulations if not across the world, at least in the European Union. Common regulations would help with the organisation of international projects. Would it be thinkable that a book about visual arts is published without any visual material but only with reference to free and copyright-free access content? Bearing in mind some incredible and unaffordable copyrights demanded by various institutions and copyright holders.

History of Art
We catalogue both slide and digital images for teaching purposes.
Management: it is important to have clearly agreed procedures and policies in place regarding the cataloguing of and access to digital images; Copyright: having a clear policy regarding obtaining permission to use images; Quality: preserving images as archive quality images; Integration: issues such as access and metadata are important and relate to image retrieval by the users. The problem for small digitisation projects is that they may not be fully integrated into a larger network thereby reducing the effectiveness of access. It is never too early to consider integration. Good project management, with clearly defined objectives, is important in developing and maintaining digital preservation projects. Skill training is important for staff involved in projects at a local level. Because of the nature of the work, digitisation projects may attract high staff turnover therefore ways of attracting and retaining good staff may be an issue. Digitisation projects often employ highly skilled staff for short contract durations. Once the project has ended these skills are often lost. It is of benefit if the skills of staff can be retained at the local level where possible. It is also important to train users how best to access resources and use them to their full potential.

History of art and visual culture
Quality and access; the preservation (alongside the visual reproduction) of a written record of when work was photographed and where (if relevant)

History of the 16th and 17th century
Quality, access and of course copy-right

History of Art: 19th century art particularly prints
Copyright, quality, access, metadata, uniform standards of digitising. Simple guidelines for the non-electronically experienced digitiser!

History
Early Modern England, for images now, State Papers Domestic Elizabeth in TNA. In the future putting images of all Elizabethan docs freely on line. Also creating word search for early modern spelling.
People seem to work from the point of view of the digi-expert, web designer, archive owner. The user should be considered first. After all the digital images are for use, the originals can be preserved.

History: Family History
Quality - grainy, poor or otherwise degenerated images are of little use to anyone. Access - images that are difficult to access are of similar little use but to a favoured few.

History

Local History archives (e.g. new photographs of old houses, collecting old photographs and postcards, transcribing old deeds and documents, wills, leases, manorial records, population survey data, etc)
In order of importance to me: Accessibility of centrally held data (using meta-data, e.g. as in A2A); Accessibility of locally held data (std. file naming conventions are important); Quality and authentication of the data; Copyright preservation. Sensitivity of certain specific material (e.g. house-interior photographs and personal data could be of use to criminals)

History: Social history and decorative arts
Access, copyright

ICT Manager for Community Services, Leics CC
Copyright; metadata - who/what/why/when; finding images; education of people taking pictures/scanning pictures to the correct quality; links to other systems holding more detailed data about objects. The tendency of people with digital cameras to take vast volumes of pictures - only some of which are worth preserving! The tendency of staff to think that using a scanner equates to a quality image (ie education of users/staff)

IT: Information Systems; Digital Preservation; Digital Memory
Physical preservation (low cost but RELIABLE storage, especially of high quality masters). - Descriptive metadata (from a preservation perspective, I'm assuming that other issues, especially copyright, will depend mainly from this requirement on standard and preserved descriptive metadata).

IT: I work in a library, from an IT background.
Access: what's the point of it existing if it can not be accessed? Metadata: there is no point in it existing if it can not be found. Quality sounds important, but nearly all images today are of a quality where they can be used. Open standards are important. Am I allowed to say something about content? I think that it is important to archive images of the mundane and everyday, as well images of event. i.e. average streets/homes/offices/etc today.

Image retrieval
Without quality the images are of limited value. IF there are copyright issues then there are potentially limits on exploitation. Quality metadata is expensive to produce yet crucial to the adequate preservation and re-use of images. Given my answer to Q1, it is not surprising that I would answer image retrieval. There has been considerable work on image information seeking behaviour in the last decade and also work on CBIR. The gap between what CBIR can achieve and what users require remains huge. Continued digitisation of images, creation of millions of images worldwide is of limited value if it is not matched by systems which enable intellectual content to those images. I have no axe to grind in favour of CBIR or of concept-based retrieval or a hybrid of the two. I do feel strongly that without effective retrieval methods which meet the needs of users there remains the question why on earth are we creating so many digital images.

Librarian
As a librarian, being able to trace and re-use digital images, to reduce damage to original prints or glass slides, and to promote our images collections - especially re. fund-raising.
copyright - very difficult to trace in archive collections donated 50 or more years ago; metadata - choice of schema; storage formats - and need to migrate data to future

standard formats; storage space (especially for Tiff and raw data files) - location and cost of this;
Image management systems/database systems - currently cost prohibitive for small collections. Related catalogue and accessions record data - e.g. if an image is part of a larger collection, the provenance and subject descriptions of the larger collection are very important to preserve with the image. Hierarchical catalogues and databases are required. The cost of current storage space, but also the cost of future migration to as yet unknown standards and formats.

Linguistics: endangered languages
data management (including metadata) as an ongoing process. The most important issues for the preservation will depend on the nature of an image's content and user community. Good data management is especially important when there is variety in the needs of the user communities.

Literary studies
Access, quality, copyright. How much more artwork can be digitised and made available in that form

Literary studies: Digitising works on paper, postcards, prints, drawings etc.
Storing metadata for each digital image is important plus having a preservation strategy for your digital assets. There seems to be so much conflicting advise in this area. Plus knowing what level of resolution to digitise your collections whether for publication or long term preservation is also something that should be considered from the outset before embarking on any major digitisation project. Just having a long term digital preservation strategy is the most important part of the process. It would be helpful if there was more information on producing and updating these kind of strategies focused around image collections.

Literary studies: Manuscripts and rare books
Quality, access, metadata, data migration, description. Reduplication, shared resources, user need

Literature, so photos (people, places) and photos of manuscripts
copyright, access, quality in that order; not sure what metadata means.

Literature
The archive of Osip Mandelstam (1891-1938), one of the greatest poets of the C20th. His manuscripts and associated material are scattered over the world, and the only realistic prospect of reuniting them is in virtual form.
Free universal access is paramount. Preservation of images for material such as ours, some of which is approaching 100 years old and beginning to fade and disintegrate, is also a matter of the greatest importance. Metadata (scholarly apparatus and existing texts for comparative purposes in our case) is the icing on the cake.

Medicine: History of Medicine
Accessibility and ensuring the electronic form the images are transferred to is a stable one.

Moving image
1. Preserving them at all (it exists) 2. Cataloguing/metadata (what it is) 3. Rights-related access restrictions (who can see it). Should we constantly migrate old image data to new formats, or emulate the display capabilities of old parsers? Is it more

important to consider what might be an enduring format, or to consider how modern machines might render old data?

Museums: Digital images as they relate to museums
As a small museum, we have a collection of digital images of our objects. Understanding and implementing copyright issues for our images. Also, ensuring that when technologies advance, our images are still available for us to use.

Museums: images of items in museum & archive collections - artworks, printed items, museum objects (all disciplines), manuscripts, etc...
A standard framework for internal image file descriptive metadata and proper application support (even if it means a very restricted list of recommended applications for manipulating archival images). Straightforward, workable copyright compliance procedures. Simple guidelines about quality standards to guide digitisation projects. I feel very strongly that there should be a standard internal descriptive metadata standard (like EXIF) across designated file formats and 'persistent', in that good quality image editing applications (if not all) will not casually destroy the data. It would be helpful if someone could draw together the state of play in this particular area and offer a set of guidelines.

Museum collections images
Having the processes in place in the first case to allow digital preservation to occur. Metadata is of interest in particular as there are bewildering number of standards available and it is often difficult to know how to choose the best tool for the job. General and absolute minimum standards which can be built upon by organisations that have different (metadata) requirements would be a helpful step towards removing some of the guess work from this problem. I feel that there should be more concentration metadata for resource discovery in these standards. A discoverable object/image is far more likely to survive if it is easily accessible to the general user. Finally, only that it is an active rather than a passive process - organisations that embark upon digitisation projects really need to be made acutely aware of this fact.

Museum Collections: Manage museum collections with digital record photography. Also archaeological archives that perhaps may have digital images in the future?
Actual preservation with hard ware systems changing so often; copyright: many non curatorial staff are oblivious to it.

Museum digitisation for Collections Management Systems and Web sites
To me, the thing that most concerns me about digital images is knowing how best to store images for archiving. Since CDs and DVDs are relatively new, there is the concern that despite lab testing and the '100 year' life of a CD, that this could be tragically wrong. Unlike an object where deterioration can be seen and recorded over time, the effects of using CDs to actively look for problems, will ultimately cause problems - and so becomes a vicious circle...do we actively conserve CDs, or blindly preserve CDs. The other issue is transference of images to new media, what problems will this cause, what percentage of quality will be lost in transfer, do we continue to store the old media, and what costs will this bring. I think a lot of good work has been done to help and assist people working with this sort of media - but the biggest problem is costs. When I started my project, the budgets had already been set and funding secured, but no-one had thought about the implications of digital image preservation. No-one had looked at file sizes and how many disks would be needed to store the quantity of images we would be collecting, and no-one had even considered that there was a big difference in CD quality...I was told when I asked about this (to a senior curator) that it was ok, as we could purchase a 'bundle'

pack of CDs from a shop for about £10! The concern therefore is: How many museums that have taken the leap of faith into digitisation projects have not considered long term preservation of their images - not forgetting that for some objects, this could be their only form of reference in years to come - and how can this problem be covered to ensure that when people apply for funding, it is a prerequisite of the application stages, that this is covered and real costs noted. I know this is all quite new to museums, and probably a system will come along that will make it quite easy and safe to do...but the problem is the present and the past, not the future!

Music: (Scottish) music
The order of importance (of issues) will depend on the image and on the agency storing and disseminating the images. The (future) availability of software and hardware that can read the images

Music and its relation to other arts
Searchability of cataloguing databases - I find identifying and locating images the most difficult task in working with them.

Philosophy
That they are of high quality, that they are indexed sensibly, and that the copyright costs are not too high. That it should not be too difficult to get hold of them.

Photograph Conservation
Access, metadata, long-term preservation (file formats; storage media). Creation of simple rules to ensure good long-term preservation of digital images, that are publicised as much as possible; simplification of procedures to ensure good long-term preservation of digital images

Photographic Archive
Preservation, understanding the nature of an image that doesn't actually exist in a hardcopy format, e.g. only a bunch of numbers and the ability to keep this image through time on various media, which have yet to be tested. I'm not too worried about not being able to access them as all it takes is the ability to write a program to interpret those numbers on whatever machine, with whatever display. I don't feel these images will ever be 'lost' just might be difficult to write a program to access them. I don't think the future will be as scary as everyone thinks its going to be, so I'm not too worried about images lasting just like paper documents have for the last so many centuries.

Photographic Images
Knowledge of the digital image to third parties. Access via the web to standardised repository thumbnails of images + metadata so that knowledge of the image is widely disseminated.
Other: A form of national escrow service where digital images can be left, and maintained, independent of other considerations. If, later, the owners cannot be found or the owning legal entity is defunct then the escrow agency could proceed in the national interest.

Photography: archival photographs
Metadata and standards. Storage and migration

Photography
My main area of interest is being able to preserve and organize original negatives and prints dating back to 1911 to 2005 and into the future. The images would need

to be scanned and duplicated so that the original image does not need to be pulled out and used again for use which could in turn harm it.
I definitely think accessibility is very important but it depends on what the use of the library is. For my library I definitely need other users to be able to pull up images rapidly, type in some key words and find the images with no delay. Concerns of if a backup tape was lost - having other backups or ways of retrieving the data, etc.

Planning
legal; promising

Preservation of all archive and library materials in whatever medium
Metadata - without that they are insufficiently documented to ensure context. Other issues include the question of awareness of fragility of the medium; too few people, particularly those at the top of organisations actively encourage digitisation without thinking about the long term or the ease with digitised images can be deleted/lost/become obsolete due to technological change.

Preservation of information
Longevity of data, metadata, future of copyright. What else may happen that is not being considered now ...

Teaching of art history
Quality; access; metadata

Teaching of history of art
Quality of image; Interoperability/upgradeability so that image can still be used in the future; metadata. How to make images as accessible to as many people as possible, so that not every art history institution is busy digitising the same images.

Theatre performance
Context documentation attached - fully documented history and original context (e.g. live performance, rehearsal, publicity); full details of cast members shown in photo and whether or not (in the case of publicity photos) they performed in the actual run; photographer, artistic director etc.

Theatre, performance, live art (archives, databases, still and moving image).
Copyright remains a major headache (for once not so much cost as acquiring and keeping the various permissions/clearances and knowing they are correct). Migration and upgrading of data (as indicated in your intro) remains another major headache. E.g. We are currently migrating some databases from one university to another but seem to be faced with problems akin to reinventing the wheel. Time/costs and shortages of expertise all feature in this one.

JISC Digital Image Study

Appendix 2: List of Research/Support Bodies

Key Organisations with preservation responsibilities

Arts and Humanities Data Service (AHDS)

www.ahds.ac.uk

The Arts and Humanities Data Service (AHDS) is a UK national service aiding the discovery, creation and preservation of digital resources in and for research, teaching and learning in the arts and humanities.
The AHDS has been involved in a number of studies relating to preservation research including the JISC funded Feasibility Study into the Preservation of E-prints, the JISC funded Long-Term Retention and Reuse of e-Learning Objects and Materials study, the JISC-funded Digital Repositories Review with UKOLN, DAAT, developing a digital preservation assessment tool for use within UK HE/FE, JISC funded Digital Picture Study, a Virtual Slide Library for HEA Archaeology and PICTIVA, demonstrating long term storage and re-use of digital images. Currently, the AHDS is working on the SHERPA DP project, a two year project funded by the JISC. The purpose of this project is to create a collaborative, shared preservation environment for the SHERPA project, **http://www/sherpa.ac.uk** framed around the OAIS Reference Model. The AHDS is responsible for access to and curation and preservation of numerous large image collections arising from research and teaching in UK Higher and Further Education and other non-profit organisations. The preferred formats for digital images for the AHDS are uncompressed TIFF, PNG and SPIFF files.

Digital Curation Centre

http://www.dcc.ac.uk/

The Digital Curation Centre is jointly funded by the JISC and the e-Science Core Programme. The DCC will support expertise and practice in data curation and preservation, prompting collaboration between the Universities and the Research Councils to ensure that there is continuing access to data of scholarly interest. The initial focus will be on research data, but the policy intention is to also address the preservation needs of e-learning and scholarly communication.
Aims include:

- Establish a research programme by addressing wider issues of data curation
- Nuture strong community relationships by forming and extending the Associates Network, engaging with scientific digital curators
- Development activity leading into services by testing and evaluating tools, methods, standards and policies in realistic settings and offering a repository of tools and technical information as well as a focal point for digital curators

- Achieving a ‘virtuous circle’ by feeding expertise, experience and need into its research programme on data curation and transforming research-led innovation into services that enhance productivity of practice
- Data Curation for e-science in the UK

Completed report http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf.

Digital Preservation Coalition

http://www.dpconline.org/graphics/index.html

The Digital Preservation Coalition (DPC) was established in 2001 to foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK and to work with others internationally to secure our global digital memory and knowledge base. DPC has grown to a membership of 27 organisations, with international links to Australia, the USA, and Europe. It has been particularly important in its advocacy role and raising the profile of digital preservation in the national media in the UK. The DPC was not a JISC project, but JISC, as one of the founding members, provided staffing support to launch and develop the DPC until the organisation was sufficiently established to recruit its own staff.

The Preservation Management of Digital Material Handbook was first compiled by Neil Beagrie and Maggie Jones and is now maintained and updated by the DPC. The handbook provides an internationally authoritative and practical guide to the subject of managing digital resources over time and the issues in sustaining access to them. This is kept updated and can be accessed through their website.

Higher Education Data Service (HEDS)

http://heds.herts.ac.uk/

Funded by JISC, HEDS offers its consultancy and production services to not-for-profit organisations from any country. The consultancy services cover feasibility studies, designing digitisation units, digital management and tendering. HEDS’ work is divided between advice, production and consultancy.

Humanities Advanced Technology and Information Institute (HATII)

http://www.hatii.arts.gla.ac.uk/

HATII has continued to promote subject-based computing by building on nearly two decades of pioneering initiatives and experience in the Faculty of Arts. We have developed a cutting edge research programme in humanities computing, digitisation, digital curation and preservation, and archives and records management. HATII conducts a range of interdisciplinary research in humanities computing. Research is conducted in four main areas:

- The relationship between digital and analogue objects
- Digital creation and storage
- User evaluation
- Information retrieval

HATII were responsible for the 2002 publication of The NINCH Guide to Good Practice in the Digital Representation & Management of Cultural Heritage Materials (**http://www.nyu.edu/its/humanities/ninchguide/**) and several reports, including, **Image Digitisation Management Models - An Assessment of the JIDI Programme** and **Digital Archaeology: Rescuing Neglected and Damaged Data Resources**. Currently, HATII are involved with the work of the:

- Digital Curation Centre (discussed below)
- The DELOS network
  The DELOS network intends to conduct a join programme of activities aimed at integrating and coordinating the ongoing research activities of the major European teams working in Digital Library with the goal of developing the next generation of Digital Library technologies
- ERPANET (discussed below)
- DigiCULT
  DigiCULT is an IST Support Measure that establishes a regular technology watch for cultural and scientific heritage. Funded by the EC, DigiCULT builds on the knowledge and expertise of over 50 cultural heritage experts in order to discuss and analyse current and future trends in several technology domains that have been identified as key areas of interest to cultural heritage institutions.

Technical Advisory Service for Images (TASI)

http://www.tasi.ac.uk/

TASI emphasises that ‘digital preservation’ is both a technical and an organisational strategy. TASI advises that technical strategies are still being debated by the digital preservation community, but that a combination of one of more might be appropriate. These strategies include technological preservation, technological emulation, data migration and data refreshing. Technically, TASI prefers PNG and SPIFF formats, but as they are still relatively unsupported formats. They advise digitisers to create TIFF files so that they can be migrated to these preferred file formats at a later date.

UKOLN

http://www.ukoln.ac.uk/

A centre of expertise in digital information management, providing advice and services to the library, information, education and cultural heritage communities. UKOLN are involved ina variety of digital library research in both UK and EC funded activities. These include technical development and support of the distributed information architecture which is providing digital library resources to UK higher and further education, development of web services in this context and research in the areas of semantic web, metadata schemas and ontologies. UKOLN also has expertise in the areas of digital preservation and Web archiving, Web standards, cross-sectoral description of collections, the development of national and institutional e-print repositories, integration of digital libraries with learning management systems and institutional portals. More recently, UKOLN is collaborating with the Grid/e-Science community through joint workshops with the National e-Science Centre and in projects linking digital library concepts with Grid-enabled applications. Some of the projects that UKOLN are involved with are discussed later in this section.

European Commission on Preservation and Access

http://www.knaw.nl/ecpa/about.html

The ECPA was established in 1994 to promote activities aimed at keeping collections in European archives and libraries accessible over time. Books, documents, photographs, films, tapes and disks are all subject to decay. The ECPA aims to raise public awareness of this issue and to impress the urgency of the situation on policy makers, funding agents, and users. The ECPA acts as a European platform for discussion and cooperation of heritage organizations in areas of preservation and access. The publications of the Commission are widely distributed to institutions throughout Europe. To promote the exchange of knowledge and experience, the ECPA organises conferences, meetings and workshops.

Research Libraries Group

http://www.rlg.org/

At the end of 1994 the Commission on Preservation and Access and RLG created the Task Force on Archiving of Digital Information, charged with investigating and recommending means to ensure "continued access indefinitely into the future of records stored in digital electronic form." In May 1996 the 21-member task force, co-chaired by Donald Waters and John Garrett, completed their final report, *Preserving Digital Information,* a milestone guide for work on the long-term retention of both born-digital and digitized materials.
Both RLG and the Council on Library and Information Resources (CLIR) have made the publication widely available. In May of this year RLG, in conjunction with the

OCLC and RLG released a comprehensive guide to core metadata for supporting the long-term preservation of digital materials. *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group* is the product of the foremost international consensus-building effort directed at preservation metadata, and it is likely to become the foundation for future work in this area. The Research Libraries Group votes upon top digital preservation projects.

Visual Resources Association

http://www.vraweb.org/diag/news.htm

The Visual Resources Association is a multi-disciplinary community of image management professionals working in educational and cultural heritage environments. The Association is committed to providing leadership in the field, developing and advocating standards, and providing educational tools and opportunities for its members. *The Digital Scene*, published by the Visual Resources Association features updates, insights and important trends for managers of digital images collections. Summaries point to additional articles and resources on the web.

JISC Activity

www.jisc.ac.uk

In 2002 'A Continuing Access and Digital Preservation Strategy for the JISC 2002-2005' (http://www.jisc.ac.uk/index.cfm?name=pres_continuing) was published. This report outlines JISC's continued commitment to digital preservation. It acknowledged that digital preservation represents a complex set of challenges, which are exceptionally difficult for institutions to address individually. National action is required, and is undertaken by a variety of services and programmes either commissioned or fully funded by JISC. Three programmes are directly responsible for work in digital preservation:

- Digital Preservation and Records Management Programme
- Supporting Digital Preservation and Asset Management in Institutions
- Digital Repositories Programme

Services with preservation responsibilities are outlined below, followed by projects attached to the above programmes.

Digital Preservation and Records Programme

This JISC programme incorporates a targeted programme, projects and core activities including:

- Developing a long-term retention strategy for digital materials of relevance to HE/FE institutions in the UK and overseeing its implementation through development programmes and projects
- Providing a UK focus for the development of advice, policies and strategies for the preservation of digital materials and management of records
- Generating support and collaborative funding from and promoting inter-working with appropriate agencies worldwide in digital preservation
- Advising JISC Committees and staff on digital preservation and records management issues
- Participating in and supporting relevant working groups and committees in digital preservation and records management
- Managing JISC's electronic and paper records

Relevant activity funded under this programme:

Digital Curation Centre (see above)

Requirements and Feasibility Study on Preservation of e-prints
http://www.jisc.ac.uk/index.cfm?name=project_eprints_pres

Carried out in 2003 by a consortium of AHDS, SHERPA (University of Nottingham) and Estonian Business Archives. The Requirements and Feasibility Study on Preservation of E-Prints provides recommendations for further research and for the development of services and tools to support the long-term preservation of UK e-print content, in the context of the JISC Information Environment (IE) and the JISC Continuing Access and Digital Preservation Strategy 2002-5 (Beagrie, 2002). Though this project directly addressed e-prints, its findings can clearly be extended, for instance, to the preservation of digital images.

The File Format Representation and Rendering Project
http://www.jisc.ac.uk/index.cfm?name=project_fileformat

The aim of the project is to support the implementation of the JISC Continuing Access and Digital Preservation Strategy 2002-5: in particular to contribute to the technical foundation for a long-term digital curation centre. The key outputs of the project would also be of great use to existing digital preservation activities within the UK. It is clear that individual digital repositories cannot realistically develop their own Representation and Rendering technologies. The cost and high degree of specialised skills required to implement this infrastructure is problematic. A project of this kind will utilise the experience of the digital preservation team at the University of Leeds in order to address these complex technical issues. The project will develop tried and tested technologies, conceived by the Cedars and CAMiLEON projects.

Supporting Digital Preservation and Asset Management in Institutions Programme

Supporting institutions in long-term digital asset management and preservation forms a central theme of JISC's **Continuing Access and Digital Preservation Strategy**. JISC will manage the programme to encourage collaboration between individual projects and related initiatives such as the new Digital Curation Centre. The following themes are addressed within these projects – institutional management support, digital preservation assessment tools and institutional repository infrastructure development. Selected activity under this programme includes:

DAAT: Digital Asset Assessment Tool
http://www.jisc.ac.uk/index.cfm?name=project_daat

The aim of this project is to develop a digital preservation assessment tool for use within the UK HE/FE and research, learning and teaching communities. The proposal will provide those responsible for managing digital resources in a variety of institutional settings, including libraries, archives, data centres, computer services and research teams, with a valuable tool for identifying the preservation needs of their digital holdings. It will do so in a way which allows scarce resources to be focussed on those assets where the risk of loss and cost of loss is greatest. This project is due for completion in May 2006.

Digital Preservation Training Programme
http://www.ulcc.ac.uk/dptp/

The Digital Preservation Training Programme (DPTP) is a project funded by JISC under its Digital Preservation and Asset Management programme, or JISC 4/04 as it is more commonly known. The project is led by ULCC, working with its partners the Digital Preservation Coalition, Cornell University and the British Library. The project's aim is to develop a modular training programme with content aimed at multiple levels of attendee. It builds on the excellent foundations of Cornell's Digital Preservation Management Workshop, whose development is funded by the National Endowment for the Humanities. It builds on existing exemplars of training and information provision, including the NEH-funded Cornell University digital preservation course, the DPC's travelling 1-day workshop, the "Preservation Management of Digital Materials" handbook, and training from existing JISC-funded services such as AHDS.

LIFE (Lifecycle Information for E-literature)
http://www.jisc.ac.uk/index.cfm?name=project_life

LIFE will explore and develop a life cycle approach to costing digital archiving for e-journals. After reviewing the existing state of knowledge, it will implement a number of methodologies to selected e-materials for the study. An international conference will be held to evaluate and validate the findings. This project will be completed in September of this year

METS Awareness Training
http://www.odl.ox.ac.uk/projects.htm

This project aims to raise general awareness of METS and other closely related emerging standards both within the Programme and among the wider community served by JISC. This project is run by the Oxford Digital Library.

Personal Archives Accessible in Digital Media (paradigm)
http://www.paradigm.ac.uk/

The exemplar strategies that this project will develop with political papers will be of use for any institution that collects, preserves, and maintains access to private papers. Such institutions are not confined to major research libraries, but include a broad sweep of institutions in HE, and in other sectors, including national libraries, museums and galleries. Each of these institutions will benefit from a project that develops best-practice guidelines rooted in practical experience in the archival and preservation aspects of digital private papers. This project is run by the Oxford Digital Library.

PRESERV (PReservation Eprint SERVices)
http://preserv.eprints.org/

Working with the National Archives, the project will link Eprints through a Web service to PRONOM software for identification and verification of file formats. The project will emphasise automation, will provide modular tools for capturing metadata and will enable the identification and verification of file formats. The project will scope a technology watch service to populate and update PRONOM where full automation is not feasible for file format recognition.

SHERPA Digital Preservation: Creating a Persistent Preservation Environment for Institutional Repositories
http://ahds.ac.uk/about/projects/sherpa-dp/

This project will create a collaborative, shared preservation environment for the SHERPA institutional repositories project framed around the Open Archiving Information Systems (OAIS) Reference Model. The project will bring together the SHERPA institutional repository systems with the preservation repository established by the Arts and Humanities Data Service to create an environment that fully addresses all the requirements of the different phases within the life cycle of digital information.

Digital Repositories Programme

The aim of the JISC Digital Repositories Programme is to bring together people and practices from across various domains (research, learning, information services, institutional policy, management and administration, records management, and so on) to ensure the maximum degree of coordination in the development of digital repositories, in terms of their technical and social (including business) aspects. Many strands of this programme have relevance to digital image preservation issues.

- **ASK: Accessing and Storing Knowledge -** will develop a suite of open source software artefacts that support learners, researchers and teachers in securely accessing and sharing learning objects.
- **CLADDIER: Citation, Location, And Deposition in Discipline and Institutional Repositories -** will build and deploy a demonstration system linking publications held in two institutional repositories (Southampton University and the CCLRC) with data holdings in the British Atmospheric Data Centre.
- **Community Dimensions of Learning Object Repositories -** will identify and analyse the factors that influence practical uptake and implementation of learning object repositories, with a focus on social and cultural issues.
- **Grade: Scoping a Geospatial Repository for Academic Deposit and Extraction -** will investigate and report on the technical and cultural issues around the reuse of geospatial data within the JISC Information Environment in the context of media-centric, informal and institutional repositories.
- **MIDESS: Management of Images in a Distributed Environment with Shared Services -** will explore the management of digitised content (especially images) in an institutional and cross-institutional context through the development of a digital repository infrastructure.
- **RepoMMan: Repository Metadata & Management -** will assist the development of repository infrastructure in several key areas by: assessing the feasibility of automated population of object metadata, conducting detailed user requirements analysis and review of associated digital rights management issues, adapting and providing a human interface to a generic workflow framework.
- **SHERPA Plus -** is a major initiative to support repository development in all UK Higher Education institutions, building on the work of the SHERPA project, and will produce: advocacy strategies and resources for the establishment of new, and further population of existing, repositories; support for policy development, and reviews and analysis of extending repository holdings with datasets, multimedia, grey literature, learning objects and other content types.

Work within the Digital Repositories Programme, but funded outside the 2005 call for proposals, includes:

- **GNU Eprints (Eprints.org).** This is free software that creates online archives. It is being developed primarily at Southampton University. Current development work aims to move key support infrastructure for Eprints into the user community. This means progressively devolving and sharing responsibility for management, code development, user support and marketing.

Two studies focusing on **community image collections (commissioned by the JISC Images Working Group)**: a study looking into the feasibility of creating a network of community based image archives:

- **Bridging the Gap – Investigating Community Led Image Collections (CLIC)**
  http://clic.oucs.ox.ac.uk

This project is lead by Jonathan Miller, University of Oxford, and is looking at the feasibility of creating a network of community-based image archives.

- A study to define a framework for deposit of sensitive and clinical recordings
  This study lead is lead by Rachel Ellaway, University of Edinburgh and will assess existing models as well before developing an agreed framework for clinical recordings depositing and access.

The **JISC Images Working Group** was established to provide advice on collecting priorities and development priorities for still image resources through a process that identifies and responds to user needs and supports the execution of the JISC Strategic Framework (incorporating the JISC Collections Strategy and the JISC Development Strategy.Other work commissioned by the JISC Images Working Group includes:

- **The Digital Picture**

  http://thedigitalpicture.ac.uk/home.html

  This study is being undertaken by the Arts and Humanities Data Service and the study will look into the use of digital pictures within arts education.

- **A Time Based Media and Image Portal Demonstrator**
  http://www.jisc.ac.uk/funding_visualandsoundportal.html

The project will take part in two phases, the first phase will be a scoping study of the functional and technical requirements of a portal to serve both medial types and the subject to the study and based on the recommendations, the second phase will be to build the portal demonstrator.

RESEARCH

DELOS (Network of Excellence on Digital Libraries) Project

http://www.delos.info/

DELOS WP6 (deliverable D6.1.1)Framework for Testbed for digital preservation experiments (Vienna University of Technology)
A Framework for a Digital Preservation Testbed
The Preservation Cluster within the DELOS Group has several major objectives in order to lay the foundation for testbeds and necessary metrics and tools for assessing preservation strategies, to raise the profile of digital preservation issues within the Digital Library Community, to collaborate with other international bodies to

ensure consistencies of digital repository standards, to ensure access to file format information and to establish the relationship between a typology of file formats and preservation strategies, to enable the definition of attributes and functionalities that need to be represented, and ensure that system development methodologies reflect preservation analysis and design issues.

Partner institutions are:
HATII, University of Glasgow (Participation led by Seamus Ross), United Kingdom
Universität zu Köln (Participation led by Manfred Thaller), Germany
Nationaal Archief Netherlands (Participation led by Hans Hofman), Netherlands
Phonogrammarchiv, Österreichischen Akademie der Wissenschaften (Participation led by Dietrich Schüller), Austria
Technische Universität Wien (Participation led by Andreas Rauber), Austria
Universita' degli Studi di Urbino Carlo Bo (Participation led by Maria Guercio), Italy
UKOLN, University of Bath (Participation led by Michael Day), United Kingdom

The digital preservation cluster of the DELOS project (www.dpc.delos.info) has published a report that provides description of a digital preservation testbed. This digital perservation cluser represents the first attempt to develop a generic framework for the evaluation of digital preservation strategies. This framework considers the object type, the preservation approach, preservation metadat and digital object attributes. As such this report describes the context of the research, the scope, the testbed environment with the sample objects and experiment and research databases, the metrics for testing, the preservation approach trials including an overview of the steps in an experiment, and finally the management of the results and products of the experiments. It is building upon the work and experiences of the Dutch Digitale Bewaring Testbed (see www.digitaleduurzaamheid.nl) and the OAIS reference model.

ERPANET

http://www.erpanet.org/

Directed by HATII, Schweizerisches Bundesarchiv, ISTBAL, Nationaal Archief van Nederland and in partnership with PADI (Preserving Access to Digital Information, at the National Library of Australia) , the ERPANET Project will establish an expandable European Consortium, which will make viable and visible information, best practice and skills development in the area of digital preservation of cultural heritage and scientific objects. Amongst ERPANET's recent workshops is an investigation (with the Digital Curation Centre) into the curation of medical databases and workflows with preservation.

Safeguarding European Photographic Images for Access (SEPIA) (part of the European Commission on Preservation and Access)

http://www.knaw.nl/ecpa/sepia/linksandlit.html

This EU Project had a digitisation and preservation strand, which was completed in 2002. The working group addressed the role of preservation in relation to digitiation. It focused on selection criteria from a preservation point of view, how to organise workflow in digitisation to meet preservation requirements.

Network of Expertise in Long-Term Storage of Digital Resources – A Digital Preservation Initiative for Germany (NESTOR)

http://www.langzeitarchivierung.de/index.php?newlang=eng

This initiative began in 2003, and was established according to the UK's Digital Preservation Coalition. Collection and preservation of digital objects is a seminal objective for NESTOR.
The project's objective is to create a network of expertise in long-term storage of digital resources for Germany. As the perspective of current and future archive users is central to the project, the emphasis is put on long-term *accessibility*. Within the project the following offers will be created: a web-based information forum, a platform for information and communication, criteria for trusted digital repositories, recommendations for certification procedures of digital repositories, recommendations for collecting guidelines and selection criteria of digital resources to be archived, guidelines and policies, the concept for a permanent organisation form of the "network of expertise in digital preservation". The long-term goal is a permanent distributed infrastructure for long-term preservation and long-term accessibility of digital resources in Germany comparable e.g. to the Digital Preservation Coalition in the UK.

Variable Media Network

http://variablemedia.net

The Variable Media Network was initiatied by the Guggenheim and is now supported and coordinated in conjunction with the Centre for Research and Documentation, Daniel Langlois Foundation. One of the tasks of the Variable Media Network is to document and preserve digital art forms. The VHM puts forward an unconventional new preservation strategy, for which it has developed a set of new *terms* that encourages artists to define their work independently from medium and in terms of behaviours, so that the work can be translated once its current medium is obsolete.

Archiving the Avant-Garde

http://www.bampfa.berkeley.edu/about_bampfa/avantgardehtml

Currently, no national or international level multi-museum projects are attempting to address strategies for documenting and preserving variable media art (including "born-digital" multimedia art and Internet art. This project, begun in 2003, aims to outline a comprehensive strategy and model for documenting and preserving variable media works, based on case studies to illustrate practical examples, but always emphasizing the generalised strategy behind the rule.

PRACTICAL RESEARCH

National Digital Information Infrastructure and Preservation Program (NDIPP)

http://www.digitalpreservation.gov/

National Digital Library Program
A pilot project, developed in the late 1990s called *American Memory* aimed to investigate practices for preserving digital content at the National Digital Library Program of the Library of Congress. Though the Program acknowledges the uncertainty of providing a long-term strategy for digital preservation, in the medium-term, they plan to '[take] steps during conversion that are likely to make migration or emulation less costly when they are needed' and to make sure the bit streams generated by the conversion process are kept alive through replication and routine refreshing supported by integrity checks.[126]
This is an extensive programme based at the Library of Congress.
http://www.digitalpreservation.gov/
The Digital Preservation Program will seek to provide a national focus on important policy, standards and technical components necessary to preserve digital content. Investments in modelling and testing various options and technical solutions will take place over several years, resulting in recommendations to the U.S. Congress about the most viable and sustainable options for long-term preservation.

'Exploring Collaborations to Harness Objects in a Digital Environment for Preservation' ECHO DEPository
http://www.ndiipp.uiuc.edu/index.php?option=com_content&task=view&id=12&Itemid=28

This project aims to address the issues regarding collecting, managing, preserving and making useful cultural digital information. This project is a collaboration between University of Illinois at Urbana-Champaign (the Library, the Graduate School of Library and Information Science, the National Center for Supercomputing Applications); the Online Computer Library Center (OCLC); the Perseus Project at Tuft's University; the Michigan State University Library; and an alliance of state libraries from Arizona, Connecticut, Illinois, North Carolina and Wisconsin.
One of the four strands of this programme includes research into 'Long-term semantic preservation research'. Using sample content collected in the project, GSLIS researchers will carry out long-range research on techniques for migrating the semantic content of documents (and document structures) across generations of

[126] Caroline R. Arms, 'Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress, *RLG DigiNews,* Vol. 4, No. 3.

encoding schemes. Advances in automated markup interpretation and inference will be applied to the problems of long-term digital preservation.
(Voted one of ten promising Digital Preservation Initiatives RGL DigiNews Vol. 9, no. 4)

Selected Digital Preservation Projects at the Koninklijke Bibliotheek - ***National library of the Netherlands***

DARE: Digital Academic Repositories

DARE aims to modernise the management of Dutch academic information by putting an infrastructure system in place and providing advanced services for the digital recording, accessing, storage and distribution of the Dutch academic output
Digital Repository Certification Task Force

The KB is a member of this international working group, set up at the initiative of RLG and NARA. Its purpose is to produce certification requirements for establishing and selecting reliable digital information repositories.

Emulation

Emulation is a preservation strategy that makes it possible to view digital objects in authentic form. The KB started a 2 year-project in April 2005 to further develop this strategy and build an operational emulator.

European Task Force Permanent Access

This Task Force was set up in 2004 after the EU Conference Permanent Access to the Records of Science. Its members want to realise a European infrastructure to secure permanent access to the digital records of science.

ICABS

IFLA-CDNL Alliance for Bibliographic Standards
The ICABS programme aims to maintain, promote, and harmonize existing standards and concepts related to bibliographic and resource control. In this context, the KB will carry out an international inquiry of international standards in digital archiving and latest achievements in the development of permanent access technology.

Tiff-archive

In November 2003 the KB started a project to investigate the possibilities of long-term storage and re-use of digitised materials of Dutch heritage institutions. In co-operation with a limited number of these institutions a pilot-system for the archiving of TIFF masters will be set up and a business model will be drawn up to explore the feasibility of a national service.
Selected Completed projects – National Library of the Netherlands

NEDLIB

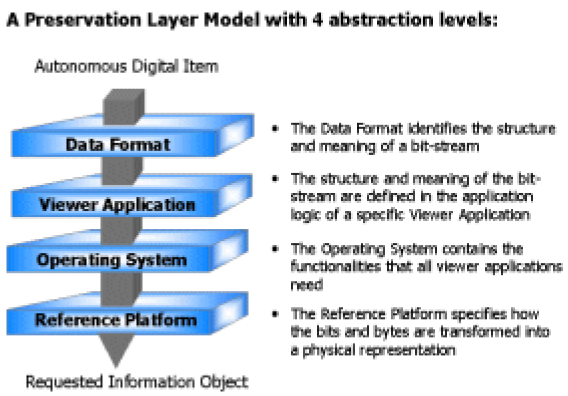
NEDLIB was a collaborative project of European national libraries. It aimed to construct the basic infrastructure upon which a networked European deposit library could be built. The objectives of NEDLIB concurred with the mission of national deposit libraries to ensure that electronic publications of the present can be used now and in the future.

PREMIS

The PREMIS working group, jointly sponsored by OCLC and RLG, was composed of international experts from institutions that had developed or were currently developing digital preservation capacity.
PREMIS membership included representatives from a variety of domains interested in digital preservation, including libraries, museums, archives, government, and the private sector.
The objectives of PREMIS were to:
Develop a core preservation metadata set, supported by a data dictionary, with broad applicability across the digital preservation community.
Identify and evaluate alternative strategies for encoding, storing, and managing preservation metadata in digital preservation systems.
Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community (September 2004) (PDF:1.25MB/66pp.)

Preservation Manager
The Preservation Manager will store information on the fileformats that are kept in the e-Depot. This is done according to a structure consisting of *View Paths* and *Preservation Layer Models* (PLM). A **View Path** consists of the software (including version numbers and required *patches*) that can be used to view a stored document. The PLM describes the different layers on which that software runs. A possible PLM could be: data format, viewer application, operating system, and hardware platform.

To ensure the rendering of stored documents, every fileformat should be linked to at least two View Paths (preferably even more). The Preservation Manager will then be able to select an alternative View Paths for a stored fileformat if technology changes cause obsolescence of a View Path. As soon as obsolescence is expected for specific parts of certain View Paths, the Preservation Manager can link this to the documents that may be at risk. For this group of documents a permanent access solution is required.



The development of the Preservation Subsystem is a joint project of the KB and IBM, as a follow-up of the project to develop the e-Depot and its technical core DIAS. A stand-alone version of the Preservation Manager has been delivered in April 2004, to be implemented in the e-Depot in the summer of 2004.

Universal Virtual Computer (UVC) for images
http://www.kb.nl/hrd/dd/dd_onderzoek/uvc_voor_images-en.html

Together with IBM Netherlands, the KB has developed a new preservation strategy, based on the Universal Virtual Computer (UVC). With the UVC it is possible to read files without adapting them and without the original hardware or software. JPEG images can now be viewed independent of changes in technology. Afterwards, the method was extended for GIF images as well. The UVC project took place between September 2003 and April 2004.
[taken from **http://www.kb.nl/hrd/dd/dd_projecten/projecten_intro-en.html**]

The International Research on Permanent Authentic Records in Electronic Systems (InterPARES)
http://www.interpares.org/

InterPARES aims at developing the theoretical and methodological knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form. The second phase of InterPARES, (which extends the work of the first phase of InterPARES One) began in 2002 and due for completion by 2006. In addition to dealing with issues of reliability and accuracy from the perspective of the entire life-cyle of records, from creation to permanent preservation. It focuses on records produced in complex digital environments in the course of artistic, scientific and e-government activities.

Department of Preservation and Collection Maintenance, Cornell University Library

http://www.library.cornell.edu/preservation/

Relevant projects include, ‘Cornell’s Digital Image Collection Project’.
This two year project aimed to develop an archiving solution to safeguard Cornell’s digital image collections (two and a half million images). The emphasis of this project was to develop a digital preservation strategy for their digital collection. The project found that current file formats (TIFF 5.0 and 6.0 ITU-T6 compressed) were acceptable and would not need to be immediately migration to new formats. However some proprietory formatting (through XDOD scanning) had to be migrated to an open Cornell Digital (CDL) format. One outcome of this project was the development of a Digital Preservation Policy Working Group. Their website also provides a digital imaging and digital preservation tutorial.

Digital Image Archive of Medieval Music (DIAMM)
http://www.diamm.ac.uk

The DIAMM project aims to digitise and enhance images of fragmentary British manuscripts of medieval polyphonic music; a rich and neglected repertory that survives mainly in fragmentary and often barely legible sources. A new permanent electronic archive of these images will be created to facilitate both the wider study of this music and its sources within the academic community, and as a security measure to assure their permanent preservation. When the current phase of the archive is completed, it will contain 2500-3000 images. Future phases are expected to expand it further.

NCSA/NARA: Scientific formats for geospatial data preservation: A Study of suitability and performance
http://www.ncsa.uiuc.edu/NARA/Sci_fmts_and_geodata_HDF.pdf

(National Center for Supercomputing Applications University of Illinois at Urbana-Champaign 2004)
This study investigates how experiences with scientific data management can help address the challenges of preserving federal geospatial collections in ways that make the data as accessible and usable as possible for current and future generations. This is a pilot study looking into cross-disciplinary technology transfer. The first phase of this project is complete, and results will be expanded upon in the second phase.

KEY PUBLICATIONS

<u>Key references and resources on preservation</u>

- *Continuing Access and Digital Preservation Strategy for the Joint Information Systems Committee (JISC)* 2002-5 http://www.jisc.ac.uk/index.cfm?name=pres_continuing
- Anne R. Kenney and Oya Y. Rieger, *Moving Theory into Practice: Digital Imaging for Libraries and Archives,* (California, Research Libraries Group, 2000).
- NINCH (2002) The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials http://www.nyu.edu/its/humanities /ninchguide
- *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* [research libraries group]
- Neil Beagrie, 2002, *Digital Preservation Coalition/AHDS Preservation Handbook* http://www.dpconline.org/graphics/handbook/
- Barbara Bültmann, Rachel Hardy, Adrienne Muir and Clara Wictor, *Digitised Content in the UK Research Library and Archives Sector: A Report to the Consortium of Research Libraries and the JISC,* April 2005 http://www.jisc.ac.uk/uploaded_documents/JISC-Digi-in-UK-v1-final.pdf

JISC Digital Image Study

Appendix 3: Major Vector and Raster Collections and Overview of JISC Collections

Major Larger Vector Image Content Creators
Vector graphic formats are generally use in specialist applications and it is often the case that individuals produce small numbers of files which they make available themselves or, particularly in the case of molecular and crystallographic data, are collected into repositories. Mapping is a major exception because this is normally carried out at a national level sponsored by governments. The major large scale content creators are concerned with geographic information and supply images as either CAD files or GIS data.

Commercial and subscription based vector image libraries

Ordnance Survey (www.ordnancesurvey.co.uk)

British Geological Survey (www.bgs.ac.uk)

ESRI (www.esri.com)

LAND INFO Worldwide Mapping, LLC (www.landinfo.com)

The GIS Data Depot (data.geocomm.com)

Vector Images: Major Online Image Collections
Vector images are rarely freely available, at least outside the United States, and are either restricted to non-commercial or *bona fide* research use or are offered for sale or for use on a licensed basis.

JISC Image Collections

UKBORDERS (edina.ed.ac.uk/ukborders)

Digimap (edina.ed.ac.uk/digimap)

MIMAS Landmap project (www.landmap.ac.uk)

MIMAS CrossFire

AHDS Archaeology (ads.ahds.ac.uk)

NERC Image Collections

Centre for Ecology and Hydrology (www.ceh.ac.uk)

British Antarctic Survey (www.antarctica.ac.uk)

Earth Observation Data Centre (ignis.neodc.rl.ac.uk)

Major Institutions

Ordnance Survey (www.ordnancesurvey.co.uk)

National Soil Resources Institute (www.silsoe.cranfield.ac.uk/nsri)

English Nature (www.english-nature.org.uk)

National Biodiversity Network (www.nbn.org.uk)

Royal Botanic Gardens, Kew: GIS Unit (www.rbgkew.org.uk/gis)

MAGIC - Multi-Agency Geographic Information for the Countryside (www.magic.gov.uk)

Northern Ireland Environment and Heritage Service (www.ehsni.gov.uk/natural/digital/ intro.shtml)

Macaulay Institute (www.macaulay.ac.uk)

Gigateway formerly askGIraffe (http://www.gigateway.co.uk/)

The Cambridge Crystallographic Data Centre (www.ccdc.cam.ac.uk)

Web Antibody Modelling (antibody.bath.ac.uk)

ChiMeraL (www.ch.ic.ac.uk/rzepa/chimeral)

National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/)

The Protein Data Bank (www.rcsb.org/pdb/) 32,727 structures

The Digital Chart of the World Data Server (www.maproom.psu.edu/dcw/)

National Geophysical Data Center (www.ngdc.noaa.gov)

Land Processes Distributed Active Archive Center (edcdaac.usgs.gov)

Commercial and subscription based image libraries

AutoCAD Blocks (www.autocadblock.com)

3D CAD Browser (www.3dcadbrowser.com)

CECO.NET (www.ceco.net)

CADsymbols.com (www.cadsymbols.com)
GIS Data Depot (data.geocomm.com)

---

3. Major Larger Raster Image Content Creators

When addressing the *major* raster image content creators, more emphasis will be placed on those creating images through UK Education funding streams, with particular attention to those funded through the JISC. However, to get some idea of the scale of image creation in the UK generally some mention should also be made of commercial image banks and libraries, who may have important preservation techniques and solutions to offer education, and also national, and to some extent regional, galleries, archives, museums and heritage organisations who contribute vast numbers of images useful to Education, and who again, may have specific preservation strategies worth considering.

In terms of sectors within education, the arts, cultural/heritage and medical arenas probably create and use the biggest number of raster images in teaching and research. Most of the major raster image creators mentioned below are therefore serving these markets.

3.1 Commercial and subscription based image creators with significant educational use in the UK.

- 4.1.1 Getty Images
- 4.1.2 Bridgeman Art Library
- 4.1.3 Artstor
- 4.1.4 Saskia
- 4.1.5 Corbis
- 4.1.6 Welcome Trust Medical Library

3.2 National Libraries, Museums, Galleries and Archives (approximate numbers of raster images created in brackets where known)

- 4.2.1 National Gallery, London
- 4.2.2 National Portrait Gallery, London (65k)
- 4.2.3 Victoria and Albert Museum (26k)
- 4.2.4 British Library
- 4.2.5 British Museum (5k)
- 4.2.6 Tate Gallery (65k)
- 4.2.7 National Galleries of Scotland
- 4.2.8 National Galleries of Wales

3.3 English Heritage

- 4.3.1 Images of England (c. 300k)
- 4.3.2 Viewfinder (c. 8k)

3.4 JISC Funding Streams with significant raster image output

- 4.4.1 JISC Image Digitisation Initiative (JIDI) (c.10 -15k images)
- 4.4.2 5/99 (c.20k ?)
- 4.4.3 X4L
- 4.4.4 digitisation call 2003 (hefce) (on-line census reports, 100k)

3.6 AHRC Funded

- 4.5.1 Resource Enhancement 2000 – present (c. 50 -100k)

3.7 Lottery/NOF

- 4.6.1 enrichUK.net (30,000)

3.8 Other Funded: e.g. British Academy, Arts Councils, Leverhulme

- 4.8.1 AXIS

3.9 Major Institutional Collections

4. Raster Images: Major Online Image Collections

Again, when addressing the *major* raster image collections, emphasis will be placed on those collections whose images were created through UK Education funding streams, with particular attention to those major collections funded through the JISC. There is a further distinction made between collections funded through JISC and delivering JISC funded content (like the AHDS) and subscription based services bought by JISC by with content funded from elsewhere (like SCRAN and Education Images On-line). Section 6 below includes a full audit of the former, i.e. those collections funded by JISC and delivering images funded by JISC. Sections 5.3 - 5.6 mention in passing significant other Image Collections with material of use to UK Education.

4.1 JISC Funded Image Collections

5.1.1 AHDS Visual Arts
5.1.2 AHDS Archaeology
5.1.3 Bristol Biomed
5.1.4 Early English Books on-line

4.2 Image Collections Subscribed to by JISC

5.2.1 Education Images On-line (EDINA/Getty Images)
5.2.2 SCRAN Resource Base
5.2.3 Documents On-line (National Archives)

4.3 National Libraries Museums Galleries and Archives

5.3.1 National gallery
http://www.nationalgallery.org.uk/

5.3.2 National Portrait gallery
http://www.npg.org.uk/live/index.asp

5.3.3 V & A access to images
http://images.vam.ac.uk/ixbin/hixclient.exe?_IXSESSION_=&submit-button=search&search-form=main/index.html

5.3.4 British Library images online
http://www.bl.uk/onlinegallery/homepage.html

5.3.5 British Museum compass
http://www.thebritishmuseum.ac.uk/compass/index.html

5.3.6 Tate
http://www.tate.org.uk/

5.3.7 National Galleries of Scotland
http://www.nationalgalleries.org/collections/

5.4 English Heritage

5.4.1 Images of England
http://www.imagesofengland.org.uk/

5.4.2 Viewfinder

http://viewfinder.english-heritage.org.uk/

4.5 Commercial and subscription based image libraries[i]

5.5.1 Bridgeman
5.5.2 Getty Images
5.5.3 Artstor
5.5.4 Saskia
5.5.5 Corbis
5.5.6 Welcome Trust Medical Library

5. Audit of JISC funded Image Collections

5.1 AHDS Visual Arts

No	Collection Title:	Institution:	Funder:	Funding Dates:	File Format	Number:
1	African and Asian Visual Artists Archive	University of East London (UEL)	JISC Image Digitisation Initiative (JIDI)	1997 - 1999	TIFF	1943
2	Artworld: Sainsbury Centre for Visual Arts	University of Norwich and The Oriental Museum, Durham	Joint Information Systems Committee (JISC) 5/99 Programme	2000-2003	TIFF	991
3	Basic Design Collection: Bretton Hall	National Arts Education Archive (Trust), Bretton Hall	JISC Image Digitisation Initiative (JIDI)	1997 - 1999	TIFF	663
4	Constance Howard Resource and Research Centre in Textiles: Material Collection	Goldsmiths College, University of London	Arts and Humanities Research Board (AHRB)	2002-2005	TIFF	375
5	Constance Howard Resource and Research Centre in Textiles: Slide Collection	Goldsmiths College, University of London	Arts and Humanities Research Board (AHRB)	2002-2005	TIFF	500
6	Corpus of Romanesque Sculpture in Britain and Ireland	Courtauld Institute of Art	British Academy, AHRB, the Henry Moore Foundation, the Marc Fitch Fund, the Heritage Council of Ireland and the Courtauld Institute of Art		TIFF	10000
7	Crafts Study Centre: Surrey Insitute of Art & Design	The Surrey Institute of Art & Design, University College (SIAD)	Joint Information Systems Committee (JISC) 5/99 Programme	2000-2003	TIFF	3768
8	Central Saint Martins: Museum & Study Collection	Central Saint Martins College of Art & Design, the London Institute	JISC Image Digitisation Initiative (JIDI)	1997 - 1999	TIFF	1467
9	Corpus Vitrearum Medii Aevi	Courtauld Institute of Art	unknown	1999-2005	TIFF	15000
10	Design Council Archive: University of Brighton	University of Brighton	JISC Image Digitisation Initiative (JIDI)	1997 - 1999	TIFF	988
11	Design Council Archive: Designing Britain	Design History Research Centre Archives at the	Joint Information Systems Committee (JISC) 5/99	2000-2003	TIFF	783

		University of Brighton	Programme			
12	The Dryden & Carr Collection: X4L RAPID	Tresham Institute of Further & Higher Education, Kettering; University College Northampton; and the Northamptonshire Libraries and Information Service	Joint Information Systems Committee (JISC) Exchange for Learning (X4L) Programme	June 2002 to June 2005	TIFF	1499
13	Design Council Slide Collection	Manchester Metropolitan University (MMU)	JISC Image Digitisation Initiative (JIDI)	1997 - 1999	TIFF	2800
14	Design Council Slide Collection	Manchester Metropolitan University (MMU)	Research Libraries Support Programme (RSLP)	2000-2003	TIFF	6377
15	Design magazine 1965-1974	London College of Communication (formerly the London College of Printing), University of the Arts London (formerly the London Institute)	Joint Information Systems Committee (JISC)	Jan 1996 - Dec 1997, Jul 1998 - Dec 1998	JPEG	2504
16	Elaine Thomas: Adopting a stance	The Surrey Institute of Art & Design, University College (SIAD)	The Surrey Institute of Art & Design, University College (SIAD)	Winter 2004	TIFF	73
17	fineart.ac.uk: Council for National Academic Awards Art Collection	Council for National Academic Awards (CNAA) Art Collection Trust	Joint Information Systems Committee (JISC)	2000-2003	TIFF	59
18	fineart.ac.uk: University of the Arts, Alumni Collection	University of the Arts London (formerly The London Institute)	Joint Information Systems Committee (JISC)	2000-2003	TIFF	17
19	fineart.ac.uk: Royal College of Art's College Collection	Royal College of Art, London	Joint Information Systems Committee (JISC)	2000-2003	TIFF	27
20	fineart.ac.uk: Norwich School of Art and Design	Norwich School of Art and Design	Joint Information Systems Committee (JISC)	2000-2003	TIFF	32
21	fineart.ac.uk: University of Brighton's Aldrich Collection	University of Brighton	Joint Information Systems Committee (JISC)	2000-2003	TIFF	28
22	fineart.ac.uk: University of Leeds, University Art Collection	University of Leeds	Joint Information Systems Committee (JISC)	2000-2003	TIFF	22
23	fineart.ac.uk: University of Ulster, Permanent Collection of Works of Art	University of Ulster	Joint Information Systems Committee (JISC)	2000-2003	TIFF	21
24	fineart.ac.uk: Glasgow School of Art	Glasgow School of Art	Joint Information Systems Committee (JISC)	2000-2003	TIFF	11
25	fineart.ac.uk: Duncan of Jordanstone College of Art and Design, Dundee	Duncan of Jordanstone College of Art and Design, Dundee	Joint Information Systems Committee (JISC)	2000-2003	TIFF	46
26	fineart.ac.uk: Slade School of Fine Art, Slade School Archive	Slade School of Fine Art, London	Joint Information Systems Committee (JISC)	2000-2003	TIFF	22
27	fineart.ac.uk:	Birmingham Institute of	Joint Information	2000-	TIFF	26

	Birmingham Institute of Art and Design, School of Art Archive Collection	Art and Design (BIAD)	Systems Committee (JISC)	2003		
28	The Woman's Library Suffrage Banners	The Women's Library, London Guildhall University	JISC Image Digitisation Initiative (JIDI)	1997 - 1999	TIFF	248
29	Halliwell Collection: Bretton Hall	National Arts Education Archive (Trust), Bretton Hall	JISC Image Digitisation Initiative (JIDI)	1997 - 1999	TIFF	247
30	Imperial War Museum Concise Art Collection	Imperial War Museum, London	unknown		JPEG	1707
31	Imperial War Museum: Posters of Conflict	Imperial War Museum and the Manchester Institute for Research and Innovation in Art and Design at Manchester Metropolitan University	Arts and Humanities Research Council (AHRC) Resource Enhancement	2003-2006	TIFF	4691
32	Imperial War Museum: Spanish Civil War Poster Collection	Imperial War Museum, London	unknown		TIFF	84
33	John Johnson Collections: Political Prints and Trades & Professions	Bodleian Library, University of Oxford	JISC Image Digitisation Initiative (JIDI)	1997 - 1999	TIFF	2373
34	London College of Fashion: College Archive	London College of Fashion, the London Institute	JISC Image Digitisation Initiative (JIDI)	1997 - 1999	TIFF	1117
35	London College of Fashion: Cordwainers College Historic Shoe Collection	London College of Fashion, the London Institute	unknown		JPEG	1233
36	London College of Fashion: The Woolmark Company	London College of Fashion, the London Institute	unknown		TIFF	2435
37	The National Inventory Research Project (NIRP)	The National Gallery, the University of Glasgow, and Birkbeck College, University of London	The Getty Grant Program, the Arts and Humanities Research Council (AHRC) Resource Enhancement scheme and the Samuel H. Kress Foundation	2004-2007	JPEG	3835
38	Public Monuments and Sculpture Association	National Recording Project (NRP)	The National Recording Project was supported by the Heritage Lottery Fund; also by the Dulverton Trust, the Henry Moore Foundation and the Pilgrim Trust, and many regional institutions	1992 ongoing	JPEG	3835
39	Russian Visual Arts: Sheffield University	University of Sheffield, University of Exeter, and The British Library	Arts and Humanities Research Board (AHRB)	2000-2003	TIFF	223
40	South Asian Diaspora Literature and Arts Archive	South Asian Diaspora Literature and Arts Archive (SALIDAA)	New Opportunities Fund (NOF)	October 2001 - 2004	TIFF	2977
41	Spellman Collection of Victorian Music Sheet Covers	Reading University Library, University of Reading	JISC Image Digitisation Initiative (JIDI)	1997 - 1999	TIFF	803

42	Fine Art Programme 2003: Surrey Institute of Art & Design	The Surrey Institute of Art & Design, University College (SIAD)	National Fine Art Education Digital Collection (fineart.ac.uk)	January - February 2003	TIFF	148
43	Textile Teaching Collection: University College for the Creative Arts	University College for the Creative Arts at Canterbury, Epsom, Farnham, Maidstone and Rochester (formerly SIAD)	Joint Information Systems Committee (JISC) 5/99 Programme	2000-2003	TIFF	1454
44	The Tim Mara Collection	The Tim Mara Archive	The Tim Mara Trust Fund		TIFF	119

5.3 AHDS Archaeology

in progress…

5.4 Bristol Biomed

No	Collection Title:	Institution:	Funder:	Funding Dates:	File Format	Number:
1	Bristol Biomed	University of Bristol	JISC 5/99	1997 - 1999	TIFF/JPEG	8000

5.5 Early English Books on-line
125,000 PDF

[i] see **www.bapla.co.uk** for full list