Technical Guidelines for Digital Cultural Content Creation Programmes

Version 1.2: Revised 18 May 2008

This document has been developed on behalf of the Minerva Project by UKOLN, University of Bath, in association with MLA The Council for Museums, Libraries & Archives.

UKOLN is funded by the Museums, Libraries and Archives Council, the Joint Information Systems Committee (JISC) of the Higher and Further Education Funding Councils, as well as by project funding from the JISC and the European Union. UKOLN also receives support from the University of Bath where it is based.

Table of Contents

1. Introduction

Throughout Europe, international, national, regional and local initiatives are investing significant public and private sector funding to enable access to a range of cultural heritage resources through digital channels. The motivations and drivers for these initiatives may vary widely: they may encompass different types of resources, address different audiences and aim to contribute to distinct social and economic objectives.

However, the various agencies supporting digitisation programmes typically share a common concern of seeking to maximise the value of their grant awards, by requiring that the content produced should be as widely useful, portable and durable as possible. These qualities are encapsulated within the notion that resources (and the mechanisms through which resources are accessed) should be 'interoperable'.

The key to such 'interoperability' is to ensure consistency of approach to the creation, management and delivery of digital resources through the effective use of standards, the rules and guidelines that codify good practice.

Digitisation programmes already recognise the value of standards, and the adoption of a shared set of technical standards and guidelines is often a first step in seeking to ensure conformity within a programme. This document seeks to provide some guidelines for the use of standards - primarily technical standards. It is intended primarily as a resource for policy-makers, and for those implementing funding programmes for the creation of digital cultural content.

1.1. The Purpose of this Document

It should be emphasised from the outset that it is not the intention of this document to impose a single prescriptive set of requirements to which all projects must conform. It would be impossible to create a single document that captured all the context-specific requirements of many different programmes, and it is recognised that different programmes will take different approaches to conformance with guidelines. Rather, this document seeks to identify those areas in which there is already commonality of approach and to provide a core around which context-specific requirements might be built. In this sense the scope and emphasis is similar to that of the EMII-DCF *Data Capture Model*, and indeed several of the recommendations in this document are based directly on those presented in that document.

As noted in that document, the usage of these guidelines cannot guarantee 'interoperability': the precise requirements for usefulness, portability and durability of digital resources will vary from programme to programme, and the form in which standards are deployed by individual projects will reflect those requirements. Further, while the guidelines provided by this document are intended to be generally applicable, each programme will operate within a context where projects are required to conform to the constraints and standards determined by many parties (institutional, programme-wide, sectoral, regional, national, international). For example, public sector funded programmes may fall within the scope of standards mandated by national governments, or it may be desirable to share data with services themselves operating within a published standards framework.

Further, even within the lifetime of a programme, the technological environment changes and standards evolve. Programmes should maintain awareness of all ongoing standards developments relevant to their operating context.

It is important that programmes also provide additional support for projects in the form of an advisory service that can offer guidance on the interpretation and implementation of standards and guidelines, and that ensures that the recommendations in those standards and guidelines are updated to reflect significant developments.

1.2. The Role of Technical Standards

The EMII-DCF *Framework Report* highlights the definition of a standard used by the British Standards Institution (BSI):

> A standard is a published specification that establishes a common language, and contains a technical specification or other precise criteria and is designed to be used consistently, as a rule, a guideline, or a definition. Standards are applied to many materials, products, methods and services. They help to make life simpler, and increase the reliability and the effectiveness of many goods and services we use.

The appropriate use of standards in digitisation can deliver the consistency that makes interoperability possible. A high level of consistency across the digital resources made available by multiple providers means that a tool or service operating across those resources needs to handle only a limited number of clearly specified formats, interfaces and protocols. In contrast, an ever-increasing number of different formats and protocols would make such development complex, costly and at best unreliable, if not impossible. In addition, the process through which standards themselves are developed means that they capture good practice based on past experience and enforce rigour in current practice.

Standards are often defined as either:

- *de jure* – formally recognized by a body responsible for setting and disseminating standards, developed usually through the common consent of a number of interested parties. An example is a standard such as the TCP/IP set of protocols, maintained by the Internet Engineering Task Force (IETF).
- *de facto* – not formally recognized by a standards body but nevertheless widely used, and recognized as a standard by its users. An example is a file format used by a software product that has a dominant or large share of the market in a particular area, such as the Adobe Portable Document Format (PDF) (although Adobe have submitted their file format to the ISO standards body for ratification as an open standard).

A further consideration is the "openness" of a standard. This can refer to a number of characteristics of a standard. The EMII-DCF *Framework Report* (http://www.emii.org/dcf.htm) for example, highlights three aspects of primary interest to the user of a standard:

- open access (to the standard itself and to documents produced during its development);
- open use (implementing the standard incurs no or little cost for IPR, through licensing, for example); and
- ongoing support driven by requirements of the user not the interests of the standard provider.

Taking the scenario above, since the specifications of the formats, interfaces and protocols used by resource providers are openly available, multiple developers can develop similar tools and services and dependency on a single tool or platform can be avoided.

Generally, the formal processes associated with the development of *de jure* standards are regarded as ensuring that such standards are genuinely "open". In these guidelines, preference is given to open standards, but in some cases industry or *de facto* standards are also considered.

1.3. The Benefits of Deploying Open Standards

Important areas for consideration include:

- **Interoperability** It is important that content can be accessed seamlessly by users, across projects and across different funding programmes. It should be possible to discover and interact with content in consistent ways, to use content easily without specialist tools, and to manage it effectively.
- **Accessibility** It is important that materials are as accessible as possible and are made publicly available using open standards and non-proprietary formats. If material is to be a widely useful resource it will be necessary to consider support for multiple language communities and ensure accessibility for citizens with a range of disabilities.
- **Preservation** It is important to secure the long-term future of materials, so that the benefit of the investment is maximised, and the cultural record is maintained in its historical continuity and media diversity.

- **Security** In a network age it is important that the identity of content and projects (and, where required, of users) is established; that intellectual property rights and privacy are protected; and that the integrity and authenticity of resources can be determined.

Failure to address these areas effectively may have serious consequences, resulting in the waste of resources by different parties:

- **Users - the citizen, the learner, the child** They will waste time and effort as they cannot readily find or use what is most appropriate to their needs, because it is not described adequately, or it is delivered in a particular way, or it requires specialist tools to exploit, or it was not captured in a usable form.
- **Information providers and managers** Their investment may be redundant and wasted as their resources fail to release their value in use, as their products reach a part only of the relevant audience, as they invest in non-standard or outmoded practices.
- **Funding agencies** They have to pay for redundant, fragmented effort, for the unnecessary repetition of learning processes, for projects that operate less efficiently than they should and deploy techniques that are less than optimal, for content that fails to meet user needs or does not meet market requirements.
- **Creators, authors** Their legacy to the future may be lost.

However the selection of appropriate open standards is not always easy. Open standards may potentially fail to deliver their potential for a variety of reasons including:

- **Complexity** The process for developing standards may result in over-complex standards being developed, which may be difficult to exploit.
- **Failure to gain marketplace acceptance** Standards may fail to gain acceptance in the market place, with a failure to develop significant numbers of tools which exploit the standard.
- **Costs** The costs of deploying open standards may be too expensive for their intended use.
- **Lack of user interest** Users may fail to make use of services which are based on open standards, preferring to continue to use services based on proprietary solutions.
- **Enhancement to proprietary solutions** The owners of proprietary formats may respond to the challenges posed by open standards by making their formats more open, reducing the costs associated with their use of enhancing the functionality of their formats.

A paper on "*A Contextual Framework For Standards*" expands on the challenges in selecting open standards and describes a framework to assist in the selection of appropriate open standards. See <http://www.ukoln.ac.uk/web-focus/papers/e-gov-workshop-2006/>.

In the light of these issues the approach which needs to be taken by cultural heritage organisations can be summarised as "**institutions *should* seek to make use of open standards and *must* have policies and procedures in place which documents their approaches and decisions related to their use of open or closed standards**".

1.4. The Life Cycle Approach

This structure of this document reflects a 'life cycle' approach to the digitisation process, and (with some modifications) parallels the structure of the *Good Practice Handbook* developed in the Minerva project.

The document is divided into the following main sections, each reflecting a stage in that life cycle. In practice, there are relationships and dependencies between activities within these different stages and indeed some of the stages may not be strictly sequential.

1. Preparation for Digitisation
2. Handling of Originals

3. The Digitisation Process
4. Storage and Preservation of the Digital Master Material
5. Metadata Creation and Capture
6. Publication
7. Disclosure
8. Use of resources
9. Reuse and Re-purposing
10. Intellectual Property and Copyright

1.5. Requirement Levels

The approaches taken to conformance to standards and guidelines vary between programmes, along a spectrum from encouraging the adoption of good practice to mandating conformance to standards as a condition of grant award. Typically the standards and guidelines adopted by programmes encompass different levels of requirement, and it is possible to distinguish between:

- **Requirements**: Standards that are widely accepted and already in current use. Projects **must** implement standards that are identified as requirements.
- **Guidance** that represents good practice but for which there may be reasons not to treat it as an absolute requirement, for example, because those standards are still in development. Projects **should** maintain and demonstrate awareness of these standards and their potential applications.

The distinction between requirements and guidance is typically made within the context of a particular programme and the intention here is to provide a foundation document for use within many different programmes.

Within the context of the standards and guidelines for a specific programme, however, the authors of guidelines for the use of technical standards **should** distinguish clearly between requirements (if any) and guidance.

Further, in standards documents, the key words '**must**, **should** and **may**' when printed in bold text are used to convey precise meanings about requirement levels:

- **Must**: This word indicates absolute technical requirement with which all projects must comply.
- **Should**: This word indicates that there may be valid reasons not to treat this point of guidance as an absolute requirement, but the full implications need to be understood and the case carefully weighed before it is disregarded. '**Should'** has been used in conjunction with technical standards that are likely to become widely implemented during the lifetime of the project but currently are still gaining widespread use.
- **May**: This word indicates that the topic deserves attention, but projects are not bound by this advice. '**May'** has therefore been used to refer to standards that are currently still being developed.

This vocabulary is based on terminology used in Internet Engineering Task Force (IETF) documentation.

Those key words are used in the remainder of this document. Within the context of the standards and guidelines for a specific programme, the authors **should** adapt the requirement levels specified in this document to those of their own contexts; authors **should** make appropriate use of these key word conventions to convey this.

IETF RFC 2119 Key words for use in RFCs to Indicate Requirement Levels <http://www.ietf.org/rfc/rfc2119.txt>.

1.6. Summary

This document has sought to provide a core set of guidelines, rather than to attempt to reflect the different requirements of many different programmes and projects. The implementers of digitisation programmes and projects will need to adapt these guidelines to the specific contexts in which they are operating, to select, to customise and to supplement as required. However, it is hoped that as a core, they can provide a starting point that is useful in many different contexts.

1.7. References In This Document

This document seeks to provide links to the definitive sources of reference for the standards described in the document. In addition links to further information, which try to provide a more readable context to the standards, are also provided. This additional guidance material may be reports which have been commissioned to document standards and best practices or entries in Wikipedia. It should be noted that readers should be aware of possible errors, biases or misleading information which made be present in the guidance materials.

1.8. About This Version

This version of the guidelines maintains the structure and approach taken in the first edition. The references given in the document have been checked and out-dated references have been updated or removed. Additional new sections have been added, including a section on Web 2.0.

2. Project Planning

2.1. Project Definition and Attributes

Firstly it is helpful to try to define exactly what a project is - what makes it different from other activities such as running an inter-library loan service or maintaining an online catalogue, and we will look at the distinguishing attributes of a project.

A project is an activity which achieves specific objectives through a set of defining tasks and effective use of resources.

Projects have a number of distinctive attributes. The specific project objectives can be grouped under three general headings: **quality** (which we can define as fitness for purpose or specification level), **costs** i.e. the budget, and **time** (to completion). Each project will have some key objectives which tend to be more important than the others e.g. the quality of the digital images is paramount and the purchase of (relatively) expensive scanning hardware may be acceptable in order to achieve this aim.

In any project, people are a fundamental key to success and provide the links which facilitate the achievement of the project objectives. This is shown in Figure 1 as a "Triangle of Objectives".

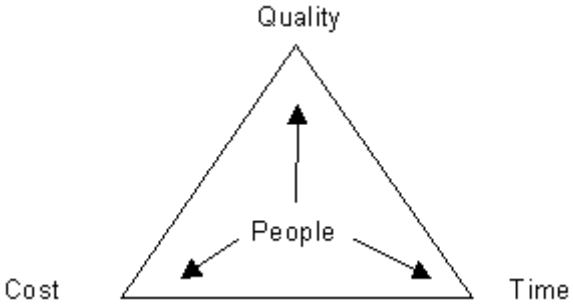

Figure 1: Triangle of Objectives

Objectives should be SMART:

Specific: expressed singularly

Measurable: ideally in quantitative terms

Acceptable: to stakeholders

Realistic: in terms of achievement

Time-bound: a timeframe is stated

Projects have a defined time frame or finite life span with a beginning and an end. Unlike the example of maintaining a service, such as an online catalogue which is an on-going and continuous activity, a digitisation project will have agreed start and end dates, which are frequently determined by the availability of project funding and/or the availability of project staff (or others) to carry out the work.

Projects are often unique either because they have never before been attempted or because the mix of parameters is customised for the particular activity.

2.2. Project Life Cycle

Projects can be described by using a life cycle approach and the four phases of the life cycle are shown below in Figure 2.

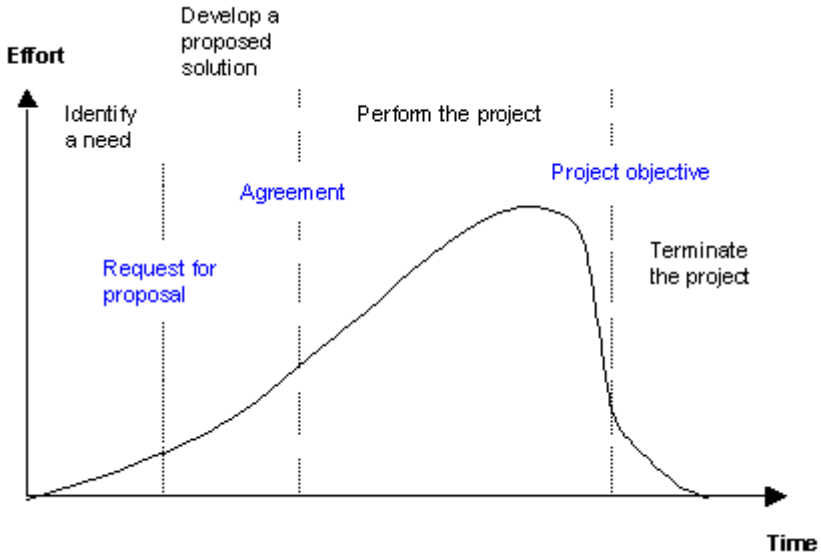

Figure 2: Project Life Cycle

In the first phase, a need is identified by the client, customer or funder. This process should result in a Request for Proposals (RfP) which describes and defines the needs and requirements. We can call this phase Initiation.

The second phase is characterised by the development of proposed solutions and the Bidding process.

The third phase in which the project is executed covers detailed planning and Implementation.

The final phase is terminating the project or Closure. In some cases this is marked with formal acceptance by the customer or client with signed documentation.

There are two important additional activities associated with projects that are worthy of special mention here: evaluation and dissemination. It should be noted that both evaluation and dissemination are not confined to the later phases of the project. The process of liaising with users and stakeholders, gaining feedback and facilitating interaction should begin as early as possible. Similarly, the mechanisms for disseminating information about project activities must begin at project start-up and continue throughout the duration of the work.

The remainder of this section concentrates on the key activities of the third phase Implementation, and covers planning, monitoring, and controlling.

2.3. Why Planning is Important

In order to achieve the objectives of any project it is essential to look at the details of the work required, which includes identifying specific tasks and estimating time to complete them, estimating associated costs, identifying who will perform the tasks and highlighting areas of risk together with devising appropriate contingency plans. In some cases, an outline of this information is required in the bid for funding.

It is usually part of the responsibility of the project manager to create the project plan and to update it on a regular and frequent basis. This is an important point - project plans are not made in tablets of stone! They are dynamic and must reflect the current situation. In most projects there are a number of "unexpected" challenges or events which may affect the timescales, costs and outcomes of the project. With good planning these unexpected events can be dealt with effectively and will not cause insoluble difficulties to the project team.

There are a number of tools available to assist with the planning process and these are described in more detail below. Which tools are used will depend on the size and nature of the project.

2.4. Formal Methods of Project Planning and Software Tools

There are a number of formal methods for managing projects. One such example is PRINCE 2. PRINCE is an acronym for PRojects IN Controlled Environments and is a complete methodology for managing all types of projects that is widely used by government departments, public bodies and the private sector in the UK. There is growing international use of the method.

Many project planning software applications are available which can help to facilitate project planning processes. Which package you use (if any) will depend on a number of factors such as size of the project, organisation preferences, licences available and previous project experiences.

Guidance:

PRINCE 2, OGC,
<http://www.ogc.gov.uk/methods_prince_2.asp>
Available 2008-05-08

Project planning, Wikipedia,
<http://en.wikipedia.org/wiki/Project_planning>
Available 2008-04-28

Project Management for a Digitisation Project, TASI,
<http://www.tasi.ac.uk/advice/managing/manage.html>
Available 2008-04-28

Managing the Project, TASI,
<http://www.tasi.ac.uk/advice/managing/project.html>
Available 2008-04-29

Project management, Good Practice Guide for Developers of Cultural Heritage Web Services, UKOLN,
<http://www.ukoln.ac.uk/interop-focus/gpg/ProjectManagement/>
Available 2008-04-28

2.5. Staff Training and Development

Projects will need to ensure that staff involved in development activities have received adequate training in order for them to complete their tasks to an adequate level. Organisations should have appropriate training and development strategies in place.

Staff training and development will cover a range of areas, depending on the nature of the development work. However these areas may well include: maintenance and preparation of materials for digitisation; digitising activities; use of hardware and software; creation and management of metadata; information flow and other workflow activities; legal issues and management issues.

Guidance:

Handbook on Cost Reduction in Digitisation, section 3.3, MINERVA Plus,
<http://www.minervaeurope.org/publications/CostReductioninDigitisation_v1_0610.pdf>
Available 2008-05-08

2.6. Disaster Planning

When developing your project plan you should also address the need for disaster planning. Disaster planning might include the identification of risks (which may be documented in a risk register), weighting the probably of risks and developing contingency plans.

Guidance:

Emergency management, Wikipedia,
<http://en.wikipedia.org/wiki/Disaster_planning>
Available 2008-04-28

3. Preparation for Digitisation

3.1. Background

Projects must develop a good knowledge of the collections to be digitised and the uses to be made of the digital resources created. Projects **should** be aware of large-scale digitization initiatives and methods for cost reduction such as outsourcing, automating digitisation and metadata creation, streamlining workflow, continuous improvement and quality assurance.

Every physical object **should** be catalogued before being digitised. If objects selected for digitisation have not been catalogued, then this should be done during the project.

Cataloguing is important for:

- Knowledge about and interpretation of the object to be digitised, also for preservation purposes.
- Contextualisation of the object; the catalogue links the object with the collection or family of objects it belongs to.
- Finding and understanding of the original object and of the digital resource representing it.

When selecting digitisation hardware and software, projects **must** take into account characteristics of the originals such as format, size, condition and the importance of capturing accurately attributes such as colour.

Guidance:

Creating Digital Images, TASI,
<http://www.tasi.ac.uk/advice/creating/creating.html>
Available 2008-04-28

The Digitisation Process, UKOLN,
<http://www.ukoln.ac.uk/nof/support/help/papers/digitisation.htm>
Available 2008-04-28

Handbook on Cost Reduction in Digitisation, MINERVA,
<http://www.minervaeurope.org/publications/CostReductioninDigitisation_v1_0610.pdf>
Available 2008-05-01

Preservation in the Age of Large-Scale Digitization, Olga Y Riegar, CLIC,
<http://www.clir.org/pubs/abstract/pub141abst.html>
Available 2008-05-01

3.2. Hardware

This document does not provide specific advice on the choice of digitisation equipment. Projects **must** demonstrate an awareness of the range of equipment available, the factors that determine its suitability for use with different types of physical object, and the ways in which it connects with other hardware such as a PC.

Projects **must** ensure that equipment selected generates digital objects of a quality that meets the requirements of their expected uses, within acceptable constraints of cost.

Project **should** seek appropriate advice before purchasing digitisation equipment or contracting digitisation services and **should** carry out an accurate costing based on the specific requirements of the project.

The choice of the hardware should take into account the physical characteristics of the physical objects: size, brittleness, presence of seals, illumination or precious bindings (e.g. cradle book scanners for manuscripts with special binding).

Performances and other characteristics declared by vendors may be checked through reference test charts.

3.3. Software

This document does not provide specific advice on the choice of software for use in digitisation. Projects **must** demonstrate an awareness of the use of software in image capture and image processing, and the hardware and software requirements of individual software products.

Project **must** ensure that software provides the functionality required given the intended uses of the digital objects created, within acceptable constraints of cost, and that software is usable by the relevant project staff.

Open source software should be evaluated along with proprietary software packages. Consideration should be given to the European IDABC programme, national laws or directives promoting the adoption of open source software by public administration (e.g. in countries including Italy, Germany and France). The need to understand the potential benefits of open source software should be balanced against potential risks in using open source software which is poorly documented, does not have a sustainable development community, etc.

Guidance:

Open Source Observatory, IDABC,
<http://ec.europa.eu/idabc/en/chapter/452>
Available 2008-04-28

OSS Watch Briefing Document, JISC OSS Watch,
<http://www.oss-watch.ac.uk/resources/fulllist.xml>
Available 2008-04-28

Free & Open Source Software Portal, UNESCO,
<http://www.unesco.org/cgi-bin/webworld/portal_freesoftware/cgi/page.cgi?d=1>
Available 2008-04-28

3.4. Environment

Establishing an appropriate environment for the digitisation process is important in ensuring that that process is effective in creating usable digital resources and ensuring that any damaging effect on the physical source materials is minimised.

Digitisation may be carried out in-house on specially purchased or existing equipment, or it may be delegated to an external agency. Project **must** understand the factors involved in this choice, not only in terms of the costs but also of the requirements for the handling of physical materials and the generation of digital objects.

4. Digitisation

4.1. The Digitisation Process

Digitisation is the conversion of analogue materials into a digital format for use by software, and decisions made at the time of digitisation have a fundamental impact on the manageability, accessibility and viability of the resources created.

It is difficult to specify fully standards for initial data capture, as requirements change over time and different resource types may have quite different requirements. However, projects **must** demonstrate that they have considered the implications of the following issues:

- the selection of materials for digitisation
- the physical preparation of materials for digitisation
- the digitisation process

4.2. Selection of Materials for Digitisation

Digitisation is the conversion of analogue materials into a digital format for use by software, and decisions made at the time of digitisation have a fundamental impact on the manageability, accessibility and viability of the resources created.

Guidance:

Deciding to Digitise, TASI,
<http://www.tasi.ac.uk/advice/managing/decide.html>
Available 2008-04-28

Guidance for selecting materials for digitisation, Joint RLG and NPO Preservation Conference, Guidelines for Digital Imaging,
<http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?fileid=0000073852:000007642156&reqid=18557>
Available 2008-04-28

Selection Guidelines for Preservation, Joint RLG and NPO Preservation Conference, Guidelines for Digital Imaging,
<http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?fileid=0000073852:000007634157&reqid=18557>
Available 2008-04-28

Handbook on Cost reduction in digitisation, Minerva Plus,
<http://www.minervaeurope.org/publications/costreduction.htm>
Available 2008-04-28

4.3. Appropriate Movement and Manipulation of Original Material

Preservation concerns apply both to the information object being digitised and to the surrogate digital object when it has been created. Those responsible for the project **must** weigh-up the risks of exposing original material to any digitisation process, especially where the items are unique, valuable or fragile, and **must** discuss the process with those responsible for the care of the originals.

FIAF has developed guidelines which provide a code of ethics for staff involved in guardianship of “the world's moving image heritage”. These guidelines may prove useful for those involved in the manipulation of other formats of cultural heritage resources.

The importance of appropriate qualification and skills should be recognised. Staff training may be required for various project phases including:

- Handling of originals.
- Technologies adopted (hardware and software).
- Cataloguing of objects.
- Management of project phases.

Guidance:

FIAF Code of Ethics,
<http://www.fiafnet.org/uk/members/ethics.cfm>
Available 2008-04-28

Preparation of Materials for Digitization, Joint RLG and NPO Preservation Conference, Guidelines for Digital Imaging,
<http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?fileid=0000073852:000007636978&reqid=18557>
Available 2008-04-28

4.4. The Digitization Process

The Technical Advisory Service for Images (TASI) provides further guidance on the digitisation process. In addition a series of resources were developed by the Arts and Humanities Data Service (AHDS). Although the AHDS no longer exist in its original form, the resources it developed are still available.

A variety of guidance regarding digitisation is also available in various publications. An important recent text is Anne R. Kenney and Oya Y. Rieger's, *Moving Theory into Practice: digital imaging for libraries and archives* (Research Libraries Group, 2000).

Of importance also are the RLG/NPO conference papers collected together in, *Guidelines for Digital Imaging* (National Preservation Office, 1998). In addition, OCLC have recently published *Gearing Up: Getting into the flow* on scaling up the digitization of special collections and materials about public/private mass digitisation agreements.

Guidance:

TASI,
<http://www.tasi.ac.uk/>
Available 2008-04-28

Guides to Good Practice, AHDS,
<http://ahds.ac.uk/creating/guides/>
Available 2008-04-28

Guidelines for Image Capture, Joint RLG and NPO Preservation Conference, Guidelines for Digital Imaging,
<http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?fileid=0000073852:000007634577&reqid=18557>
Available 2008-04-28

Cost reduction in digitisation, Working Draft Version 1, Minerva,
<http://www.minervaeurope.org/publications/costreduction.htm>
Available 2008-04-28

Shifting Gears: Get Into the Flow 2007,
<http://www.oclc.org/programs/publications/reports/2007-02.pdf>
Available 2008-04-23

Public/Private Mass Digitization Agreements,
<http://www.oclc.org/programs/ourwork/collectivecoll/harmonization/massdigresourcelist.htm>
Available 2008-04-28

Preservation in the Age of Large-Scale Digitization: A White Paper, Oya Y. Rieger, February, 2008. ISBN 978-1-932326-29-1
<http://www.clir.org/pubs/abstract/pub141abst.html>
Available 2008-04-28

5. Storage and Management of the Digital Master Material

Preservation issues **must** be considered an integral part of the digital creation process. Preservation will depend upon documenting all of the technological procedures that led to the creation of an object, and much critical information can – in many cases – be captured only at the point of creation.

Projects **must** consider the value in creating a fully documented high-quality 'digital master' from which all other versions (e.g. compressed versions for access via the Web) can be derived. This will help with the periodic migration of data and with the development of new products and resources.

It is important to realise that preservation is not just about choosing suitable file formats or media types. Instead, it should be seen as a fundamental management responsibility for those who own and manage digital information content, ensuring its long-term use and re-use. This depends upon a variety of factors that are outside of the digitisation process itself, e.g. things like institutional stability, continued funding and the ownership of intellectual property rights.

However, there are technical strategies that can be adopted during the digitisation process to facilitate preservation. For example, many digitisation projects have begun to adopt strategies based on the creation of metadata-rich 'digital masters'. A brief technical overview of the 'digital master' strategy is described in the information paper on the digitisation process produced for the UK NOF-digitise programme by HEDS.

Guidance:

Preservation Management of Digital Materials Handbook, Digital Preservation Coalition,
<http://www.dpconline.org/graphics/handbook/>
Available 2008-04-28

The Digitisation Process, HEDS, nof-digitise Technical Advisory Service,
<http://www.ukoln.ac.uk/nof/support/help/papers/digitisation.htm>
Available 2008-04-28

JISC Standards catalogue
<http://standards.jisc.ac.uk/catalogue/Home.phtml>
Available 2008-05-08

5.1. File Formats

Open standard formats **should** be used when creating digital resources in order to maximise access. (Note that file formats for the delivery of digital records to users is described later). The use of open file formats will help with interoperability, ensuring that resources are reusable and can be created and modified by a variety of applications. It will also help to avoid dependency on a particular supplier.

However, in some cases there may be no relevant open standards or the relevant standards may be sufficiently new that conformant tools are not widely available. In some cases therefore, the use of proprietary formats **may** be acceptable. However, where proprietary formats are used, the project **should** explore a migration strategy that will enable a transition to open standards to be made in the future.

If open standards are not used, projects **should** justify their requirement for use of proprietary formats within their proposals for funding, paying particular attention to issues of accessibility.

5.1.1. Text Capture and Storage

Character Encoding

A character encoding is an algorithm for presenting characters in digital form by mapping sequences of code numbers of characters (the integers corresponding to characters in a repertoire) into sequences of 8-bit values (bytes or octets). An application requires an indication about the character encoding used in a document in order to interpret the bytes which make up that digital object.

The character encoding used by text-based documents **should** be explicitly stated. For XML documents, the character encoding **should** usually be recorded in the encoding declaration of the XML declaration.

Standards:

The Unicode Consortium. The Unicode Standard, Version 4.0.0, defined by: *The Unicode Standard, Version 4.0* (Boston, MA, Addison-Wesley, 2003. ISBN 0-321-18578-1)
<http://www.unicode.org/versions/Unicode4.0.0/>
Available 2008-04-28

Extensible Markup Language (XML) 1.0, W3C,
<http://www.w3.org/TR/REC-xml/>
Available 2008-04-28

XHTML 1.0 The Extensible HyperText Markup Language, W3C,
<http://www.w3.org/TR/xhtml1/>
Available 2008-04-28

Guidance:

A Tutorial on Character Code Issues, Jukka Korpela,
<http://www.cs.tut.fi/~jkorpela/chars.html>
Available 2008-04-28

Character Encoding, Wikipedia,
<http://en.wikipedia.org/wiki/Character_code>
Available 2008-04-28

Document Formats

Text based content **should** be created and managed in a structured format that is suitable for generating HTML or XHTML documents for delivery.

In most cases storing text-based content in an SGML- or XML-based form conforming to a published Document Type Definition (DTD) or XML Schema will be the most appropriate option. Projects **may** choose to store such content either in plain files or within a database of some kind. All documents **should** be validated against the appropriate DTD or XML Schema.

Projects **should** display awareness of and understand the purpose of standardised formats for the encoding of texts, such as the Text Encoding Initiative (TEI), and **should** store text-based content in such formats when appropriate. Projects **may** store text-based content as HTML 4 or XHTML 1.0 (or subsequent versions). Projects **may** store text-based content in SGML or XML formats conforming to other DTDs or Schemas, but **must** provide mappings to a recognised schema.

In some instances, projects **may** choose to store text-based content using Adobe Portable Document Format (PDF). For a long time PDF has been a proprietary file format owned by Adobe that preserves the fonts, formatting, colours and graphics of the source document. PDF files are compact and can be viewed and printed with the freely available Adobe Acrobat Reader. However, the PDF format is being standardised and PDF/A is now an ISO Standard for using PDF format for the long-term archiving of electronic documents.

Standards:

ISO 8879:1986. Information Processing -- Text and Office Systems -- Standard Generalized Markup Language (SGML)

Extensible Markup Language (XML) 1.0, W3C,
<http://www.w3.org/TR/REC-xml/>
Available 2008-04-28

Text Encoding Initiative (TEI), The Text Encoding Initiative,
<http://www.tei-c.org/>
Available 2008-04-28

HTML 4.01 HyperText Markup Language, W3C,
<http://www.w3.org/TR/html401/>
Available 2008-04-28

XHTML 1.0 The Extensible HyperText Markup Language, W3C,
<http://www.w3.org/TR/xhtml1/>
Available 2008-04-28

PDF/A Competence Center,
<http://www.pdfa.org/doku.php>
Available 2008-04-28

Other references:

Portable Document Format (PDF), Adobe,
<http://www.adobe.com/products/acrobat/adobepdf.html>
Available 2008-04-28

Guidance:

AHDS Guide to Good Practice: Creating and Documenting Electronic Texts, AHDS,
<http://ota.ox.ac.uk/documents/creating/cdet/>
Available 2008-04-28

PDF/A, Wikipedia,
<http://en.wikipedia.org/wiki/PDF/A>
Available 2008-04-28

TEI, Wikipedia,
<http://en.wikipedia.org/wiki/Text_Encoding_Initiative>
Available 2008-04-28

5.1.2. Still Image Capture and Storage

Digital still images fall into two main categories: raster (or 'bit-mapped') images and vector ('object-oriented') images. Raster images take the form of a grid or matrix, with each 'picture element' (pixel) in the matrix having a unique location and an independent colour value that can be edited separately. Vector files provide a set of mathematical instructions that are used by a drawing program to construct an image.

The digitisation process will usually generate a raster image; vector images are usually created as outputs of drawing software.

Raster Images

When creating and storing raster images, two factors need to be considered: the file format and the quality parameters.

Raster images **should** usually be stored in the uncompressed form generated by the digitisation process without the application of any subsequent processing. Raster images **must** be created using one of the following formats – Tagged Image File Format (TIFF), Portable Network Graphics (PNG), Graphical Interchange Format (GIF) or JPEG Still Picture Interchange File Format (JPEG/SPIFF).

There are two primary parameters to be considered:

- **Spatial resolution:** The frequency at which samples of the original are taken by the capture device, expressed as a number of samples per inch (spi), or more commonly just as pixels per inch (ppi) in the resulting digital image.
- **Colour resolution (bit depth):** The number of colours (or levels of brightness) available to represent different colours (or shades of grey) in the original, expressed in terms of the number of bits available to represent colour information, e.g. a colour resolution of 8 bits means 256 different colours are available.

In general photographic images **should** be created as TIFF images.

The selection of quality parameters required to capture a useful image of an item is determined by the size of the original, the amount of detail in the original and the intended uses of the digital image. Digitising a 35mm transparency will require a higher resolution than a 6x4 print because it is smaller and more detailed; if a required use of an image of a watercolour is the capacity to analyse fine details of brushstrokes, then that requires a higher resolution than that required to simply display the picture as a whole on a screen.

Images **should** be created at the highest suitable resolution and bit depth that is both affordable and practical given the intended uses of the images, and each project **must** identify the minimum level of quality and information density it requires.

As a guide, a resolution of 600 dots per inch (dpi) and a bit depth of 24-bit colour or 8-bit greyscale **should** be considered for photographic prints. A resolution of 2400 dpi **should** be considered for 35 mm slides to capture the increased density of information.

In some cases, for example when using cheaper digital cameras, it **may** be appropriate to store images in JPEG/SPIFF format as an alternative to TIFF. This will result in smaller, but lower quality images. Such images may be appropriate for displaying photographs of events etc. on a Web site but it is not suggested that such cameras are used for the large-scale digitisation of content.

Standards:

Tagged Image File Format (TIFF), ITU International Telecommunication Union,
<http://www.itu.int/itudoc/itu-t/com16/tiff-fx/docs/tiff6.pdf>
Available 2008-04-28

Joint Photographic Expert Group (JPEG), W3C,
<http://www.w3.org/Graphics/JPEG/>
Available 2008-04-28

JPEG Still Picture Interchange File Format (SPIFF), JPEG Committee,
<http://www.jpeg.org/public/spiff.pdf>
Available 2008-04-28

Guidance:

Creating Digital Images, TASI,
<http://www.tasi.ac.uk/advice/creating/creating.html>
Available 2008-04-28

JPEG, Wikipedia,
<http://en.wikipedia.org/wiki/JPEG>
Available 2008-04-28

Tagged Image File Format, Wikipedia,
<http://en.wikipedia.org/wiki/Tagged_Image_File_Format>
Available 2008-04-28

Graphic Non-vector Images

Computer-generated images such as logos, icons and line drawings **should** normally be created as PNG or GIF images at a resolution of 72 dpi. (Note that images resulting from the digitisation of physical line drawings should be managed as described in the previous section.)

Standards:

Portable Network Graphics (PNG), W3C,
<http://www.w3.org/TR/PNG/>
Available 2008-04-28

Guidance:

Portable Network Graphics, Wikipedia,
<http://en.wikipedia.org/wiki/Portable_Network_Graphics>
Available 2008-04-28

Vector Images

Vector images consist of multiple geometric objects (lines, ellipses, polygons, and other shapes) constructed through a sequence of commands or mathematical statements to plot lines and shapes. Vector graphics **should** be created and stored using an open format such as Scalable Vector Graphics (SVG), an XML language for describing such graphics. SVG drawings can be interactive and dynamic, and are scalable to different screen display and printer resolutions.

Use of the proprietary Macromedia SWF format (sometimes referred to as Flash) may also be appropriate, however projects **should** explore a migration strategy so that they can move to more open formats once they become widely deployed. In addition, the use of text within the SWF format should be avoided, in order to support the development of multi-lingual versions.

Standards:

Scalable Vector Graphics (SVG)
<http://www.w3.org/TR/SVG/>
Available 2008-04-28

Other references:

File Format Specification FAQ
<http://www.adobe.com/licensing/developer/fileformat/faq/>
Available 2008-04-28

Guidance:

SWF, Wikipedia
<http://en.wikipedia.org/wiki/SWF>
Available 2008-05-18

SVG, Wikipedia
<http://en.wikipedia.org/wiki/SVG>
Available 2008-05-18

5.1.3. Video Capture and Storage

Video **should** usually be stored in the uncompressed form obtained from the recording device without the application of any subsequent processing. Video **should** be created at the highest suitable resolution, colour depth and frame rate that are both affordable and practical given its intended uses, and each project **must** identify the minimum level of quality it requires.

Video **should** be stored using the uncompressed raw AVI format, without the use of any codec, at a frame size of 720x576 pixels, a frame rate of 25 frames per second, using 24-bit colour. PAL colour encoding should be used.

Video **may** be created and stored using the appropriate MPEG format (MPEG-1, MPEG-2 or MPEG-4) or the proprietary formats Microsoft WMF, ASF or Quicktime.

Standards:

The reference website for MPEG!, MPEG.org,
<http://www.mpeg.org/>
Available 2008-04-28

MPEG, Wikipedia
<http://en.wikipedia.org/wiki/Mpeg>
Available 2008-05-18

5.1.4. Audio Capture and Storage

Audio **should** usually be stored in the uncompressed form obtained from the recording device without the application of any subsequent processing such as noise reduction. Audio **should** be created and stored as an uncompressed format such as Microsoft WAV or Apple AIFF. 24-bit stereo sound at 48/96 KHz sample rate **should** be used for master copies. This sampling rate is suggested by the Audio Engineering Society (AES) and the International Association of Sound and Audiovisual Archives (IASA).

Audio **may** be created and stored using compressed formats such as MP3, WMA, RealAudio, or Sun AU formats.

Standards:

MP3, Fraunhofer IIS,
<http://www.iis.fraunhofer.de/EN/bf/amm/projects/mp3/index.jsp>
Available 2008-04-28

The Ogg Encapsulation Format Version 0, RFC 3533,
<http://www.xiph.org/ogg/doc/rfc3533.txt>
Available 2008-04-28

The Ogg container format, Xoph.org
<http://www.xiph.org/ogg/>
Available 2008-04-28

Guidance:

Guidelines on the Production and Preservation of Digital Audio Objects, IASA Technical Committee, ISBN 87-990309-1-8
<http://www.iasa-web.org/pages/06pubs_03_new.htm>
Available 2008-04-28

The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy, IASA Technical Committee, ISBN 91-976192-0-5
<http://www.iasa-web.org/IASA_TC03/TC03_English.pdf>
Available 2008-04-28

Comparison of audio codecs, Wikipedia
<http://en.wikipedia.org/wiki/Comparison_of_audio_codecs>
Available 2008-04-28

Ogg, Wikipedia
<http://en.wikipedia.org/wiki/Ogg>
Available 2008-05-18

MP3, Wikipedia
<http://en.wikipedia.org/wiki/MP3>
Available 2008-04-28

AVI, Wikipedia
<http://en.wikipedia.org/wiki/AVI>
Available 2008-05-18

AIFF, Wikipedia
<http://en.wikipedia.org/wiki/AIFF>
Available 2008-05-18

AU, Wikipedia
<http://en.wikipedia.org/wiki/Au_file_format>
Available 2008-05-18

WMA, Wikipedia
<http://en.wikipedia.org/wiki/Windows_Media_Audio>
Available 2008-05-18

5.1.5. Multimedia

Multimedia formats can be used to provide integration of text, image, sound and video resources.

The W3C SMIL format **may** be an appropriate open standard for multimedia delivered over the Web.

The proprietary SWF format (often referred to as Flash) may be a suitable medium for multimedia.

Standards:

Synchronized Multimedia, W3C,
<http://www.w3.org/AudioVideo/>
Available 2008-04-28

Macromedia Flash File Format (SWF) Specification License, Macromedia,
<http://www.macromedia.com/software/flash/open/licensing/fileformat/>
Available 2008-04-28

Guidance:

SMIL, Wikipedia
<http://en.wikipedia.org/wiki/Synchronized_Multimedia_Integration_Language>
Available 2008-04-28

SWF, Wikipedia
<http://en.wikipedia.org/wiki/SWF>
Available 2008-04-28

5.1.6. GIS

GIS (Geographic Information Standards) standards can be used to support location-based services.

The ISO geographic information series of standards (or which there are over 40) addresses a diverse range of functions.

Standards:

ISO/TC 112
<http://www.isotc211.org/>
Available 2008-04-28

Guidance:

Geographic Information Standards, Wikipedia
<http://en.wikipedia.org/wiki/Geographic_information_system>
Available 2008-04-28

5.1.7. 3D and Virtual Worlds

3D models consist of complex geometric objects that may be created either by 3D scanning of original objects or by using 3D software to model the shapes. Scanning creates a point cloud which is then imported into 3D software. The captured shapes can be rendered with images to create realistic surfaces and other effects applied. Simple geometric objects may be often combined to create more complex shapes, such as the inside and outside of a building. 3D models are often imported into Virtual Reality applications to create compelling user environments.

CAD software is widely used to create architectural models but a wide range of 3D graphical software packages is available. Open standards are not well developed.

Virtual X3D is an ISO standard for virtual reality that has been developed by the Web 3D consortium from Virtual Reality Modelling Language (VRML) and which provides a system for

the storage, retrieval and playback of real time graphics content embedded in applications. Virtual Reality models should be created using the X3D format.

The DWG format may be an appropriate open standard for CAD files; this proprietary AutoDesk format is being promoted by the Open Design Alliance as a public standard.

Projects should be aware that with any proprietary solution there are potential costs and should explore a migration strategy that will enable a future transition to open standards to be made.

Standards:

Adobe Director 11, Adobe,
<http://www.adobe.com/uk/products/director/>
Available 2008-05-02

VRML, Web3D,
<http://www.web3d.org/x3d/specifications/vrml/ISO-IEC-14772-VRML97/>
Available 2008-05-02

Virtual X3D, Web3D,
<http://www.web3d.org/about/overview/>
Available 2008-05-02

Guidance:

CAD: a Guide to Good Practice, ADS
<http://ads.ahds.ac.uk/project/goodguides/cad/>
Available 2008-05-02

Creating and Using Virtual Reality, VADS
<http://vads.ahds.ac.uk/guides/vr_guide/>
Available 2008-05-02

Web3D Consortium
<http://www.web3d.org/>
Available 2008-05-02

Open Design Alliance,
<http://www.opendesign.com/>
Available 2008-05-02

.DWG, Wikipedia
<http://en.wikipedia.org/wiki/.dwg>
Available 2008-05-08

X3D, Wikipedia
<http://en.wikipedia.org/wiki/X3D>
Available 2008-05-08

VRML, Wikipedia
<http://en.wikipedia.org/wiki/VRML>
Available 2008-05-08

5.2. Media Choices

Different digital storage media have different software and hardware requirements for access and different media present different storage and management challenges. The threats to continued access to digital media are two-fold:

- The physical deterioration of, or damage to, the medium itself
- Technological change resulting in the obsolescence of the hardware and software infrastructure required to access the medium

The resources generated during digitisation project will typically be stored on the hard disks of one or more file servers, and also on portable storage media. At the time of writing, the most commonly used types of portable medium are magnetic tape and optical media (CD-R and DVD).

Portable media chosen **should** be of good quality and purchased from reputable brands and suppliers, and new instances should always be checked for faults. Media **should** be handled, used and stored in accordance with their suppliers' instructions.

Projects **should** consider creating copies of all their digital resources – metadata records as well as the digitised objects - on two different types of storage medium. At least one copy **should** be kept at a location other than the primary site to ensure that they are safe in the case of any disaster affecting the main site. All transfers to portable media **should** be logged.

Media **should** be refreshed (i.e. the data copied to a new instance of the same medium) on a regular cycle within the lifetime of the medium. Refreshment activity **should** be logged.

Guidance:

Preservation Management of Digital Materials
<http://www.dpconline.org/graphics/handbook/>
Available 2008-04-28

Using CD-R and DVD-R for Digital Preservation, TASI,
<http://www.tasi.ac.uk/advice/delivering/cdr-dvdr.html>
Available 2008-04-28

5.3. Preservation Strategies

There are three main technical approaches to digital preservation: technology preservation, technology emulation and data migration. The first two focus on the technology used to access the object, either maintaining the original hardware and software or using current technology to replicate the original environment. The work on "persistent archives" based on the articulation of the essential characteristics of the objects to be preserved may also be of interest.

Migration strategies focus on the maintaining the digital objects in a form that is accessible using current technology. In this scenario, objects are periodically transferred from one technical environment to another, newer one, while as far as possible maintaining the content, context, usability and functionality of the original. Such migrations may require the copying of the object from one medium or device to a new medium or device and/or the transformation of the object from one format to a new format. Some migrations may require only a relatively simple format transformation; a migration to a very different environment may involve a complex process with considerable design effort.

Projects **should** understand the requirements for a migration-based preservation strategy and **should** develop policies and guidelines to support its implementation.

The capture of metadata is a critical part of a migration-based preservation strategy. Metadata is required to support the management of the object and of the migration process, but furthermore, migration inevitably leads, at least in the longer term, to some changes in, or losses of, original functionality. Where this is significant to the interpretation of the object, users will rely on metadata about the migration process- and about the original object and its transformations - to provide some understanding of the functionality provided in the original technological environment.

Guidance:

Preservation Management of Digital Materials Handbook, Digital Preservation Coalition,
<http://www.dpconline.org/graphics/handbook/>
Available 2008-04-28

The State of Digital Preservation: An International Perspective
<http://www.clir.org/PUBS/reports/pub107/contents.html>
Available 2008-04-28

Planets - Preservation and Long-term Access through NETworked Services
<http://www.planets-project.eu/>
Available 2008-04-20

6. Metadata Creation/Capture

Metadata can be defined literally as "data about data," but the term is normally understood to mean structured data about resources that can be used to help support a wide range of operations on those resources. A resource may be anything that has identity, and a resource may be digital or non-digital. Operations might include, for example, disclosure and discovery, resource management (including rights management) and the long-term preservation of a resource. For a single resource different metadata may be required to support these different functions.

6.1. The Scope of the Metadata

It may be necessary to provide metadata describing several classes of resource, including

- the physical objects digitised;
- the digital objects created during the digitisation process and stored as “digital masters”;
- the digital objects derived from these “digital masters” for networked delivery to users;
- new resources created using these digital objects;
- collections of any of the above

6.2. Appropriate Standards

Metadata is sometimes classified according to the functions it is intended to support. In practice, individual metadata schemas often support multiple functions and overlap the categories below.

The curatorial communities responsible for the management of different types of resources have developed their own metadata standards to support operations on those resources. The museum community has created the SPECTRUM and CDWA standards to support the management of museum objects; the archive community has developed the ISAD(G), ISAAR(CPF) and EAD standards to provide for the administration and discovery of archival records; and the library community uses the MARC family of standards to support the representation and exchange of bibliographic metadata.

Project **should** display awareness of the requirements of community-/domain-specific metadata standards.

Projects **should** ensure that the metadata schema(s) adopted is (are) fully documented. This documentation **should** include detailed cataloguing guidelines listing the metadata elements to be used and describing how those elements are to be used to describe the types of resource created and managed by the project. Such guidelines are necessary even when a standard metadata schema is used in order to explain how that schema is to be applied in the specific context of the project.

Standards:

SPECTRUM, the UK Museum Documentation Standard, 2nd Edition

Getty Research Institute, Categories for the Description of Works of Art (CDWA)
<http://www.getty.edu/research/conducting_research/standards/cdwa/>
Available 2008-04-28

International Standard for Archival Description (General) (ISAD(G)). Second Edition.
<http://www.icacds.org.uk/eng/ISAD(G).pdf>
Available 2008-04-28

International Standard Archival Authority Record for Corporate Bodies, Persons and Families. Second Edition.
<http://www.icacds.org.uk/eng/ISAAR(CPF)2ed.pdf>
Available 2008-04-28

Encoded Archival Description (EAD)
<http://www.loc.gov/ead/>
Available 2008-04-28

Machine Readable Cataloguing (MARC): MARC 21
<http://www.loc.gov/marc/>
Available 2008-04-28

Guidance:

Online Archive of California Best Practice Guidelines for Encoded Archival Description (OAC BPG EAD), Version 2.0
<http://www.cdlib.org/inside/diglib/guidelines/bpgead/oacbpgead_v2-0.pdf>
Available 2008-04-28

6.2.1. Descriptive Metadata

Descriptive metadata is used for discovery and interpretation of the digital object.

Projects **should** show understanding of the requirements for descriptive metadata for digital objects.

To support the discovery of their resources by a wide range of other applications and services, projects **must** capture and store sufficient descriptive metadata to be able to generate a metadata description for each item using the Dublin Core Metadata Element Set (DCMES) in its simple/unqualified form. The DCMES is a very simple descriptive metadata schema, developed by a cross-disciplinary initiative and designed to support the discovery of resources from across a range of domains. It defines fifteen elements to support simple cross-domain resource discovery: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage and Rights.

This requirement does not mean that only simple DC metadata should be recorded for each item: rather, the ability to provide simple DC metadata is the minimum requirement to support resource discovery. In practice, that simple DC metadata will probably be a subset of a richer set of item-level metadata

To support discovery within the cultural heritage sector, projects **should** also consider providing a metadata description for each item conforming to the DC.Culture schema.

Projects **should** show awareness of any additional requirements for descriptive metadata, and **may** need to capture and store additional descriptive metadata to meet those requirements.

Standards:

Dublin Core Metadata Element Set, Version 1.1
<http://dublincore.org/documents/dces/>
Available 2008-04-28

DC.Culture
<http://www.minervaeurope.org/DC.Culture.htm>
Available 2008-04-28

Guidance:

Using Dublin Core
<http://dublincore.org/documents/usageguide/>
Available 2008-04-28

6.2.2. Administrative Metadata

Administrative metadata is used for managing the digital object and providing more information about its creation and any constraints governing its use. This might include

- Technical metadata, describing technical characteristics of a digital resource;
- Source metadata, describing the object from which the digital resource was produced;

- Digital provenance metadata, describing the history of the operations performed on a digital object since its creation/capture;
- Rights management metadata, describing copyright, use restrictions and license agreements that constrain the use of the resource.

Technical metadata includes information that can only be captured effectively as part of the digitisation process itself: for example, information about the nature of the source material, about the digitisation equipment used and its parameters (formats, compression types, etc.), and about the agents responsible for the digitisation process. It may be possible to generate some of this metadata from the digitisation software used.

There is, however, no single standard for this type of metadata. For images, a committee of the US National Information Standards Organization (NISO) has produced a draft data dictionary of technical metadata for digital still images.

Projects **should** show understanding of the requirements for administrative metadata for digital objects.

Projects **must** capture and store sufficient administrative metadata for the management of their digital resources.

Standards:

NISO Z39.87-2002 AIIM 20-2002 Data Dictionary -- Technical Metadata for Digital Still Images
<www.niso.org/standards/resources/Z39_87_trial_use.pdf>
Available 2008-04-28

6.2.3. Preservation Metadata

A set of sixteen basic metadata elements to support preservation was published in 1998 by a Working Group on Preservation Issues of Metadata constituted by the Research Libraries Group (RLG).

The Reference Model for an Open Archival Information System (OAIS) is an attempt to provide a high-level framework for the development and comparison of digital archives. It provides both a functional model, that outlines the operations to be undertaken by an archive, and an information model, that describes the metadata required to support those operations.

Using the OAIS model as their framework, an OCLC/RLG working group on preservation metadata has developed proposals for two components of the OAIS information model directly relevant to preservation metadata (Content Information and Preservation Description Information).

Standards

Reference Model for an Open Archival Information System (OAIS)
<http://public.ccsds.org/publications/archive/650x0b1.pdf>
Available 2008-04-28

Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects
<http://www.oclc.org/research/projects/pmwg/pm_framework.pdf>
Available 2008-04-28

Guidance:

Preservation Metadata (pre-publication draft)
<http://www.ukoln.ac.uk/metadata/publications/iylim-2003/>
Available 2008-04-28

DCC Approach to Digital Curation - under Development
<http://twiki.dcc.rl.ac.uk/bin/view/Main/DCCApproachToCuration>
Available 2008-04-28

6.2.4. Structural Metadata

Structural metadata describes the logical or physical relationships between the parts of a compound object. For example, a physical book consists of a sequence of pages. The digitisation process may generate a number of separate digital resources, perhaps one image per page, but the fact that these resources form a sequence and that sequence constitutes a composite object is clearly essential to their use and interpretation.

The Metadata Encoding and Transmission Standard (METS) provides an encoding format for descriptive, administrative and structural metadata, and is designed to support both the management of digital objects and the delivery and exchange of digital objects across systems.

The IMS Content Packaging Specification describes a means of describing the structure of and organising composite learning resources.

Projects **should** show understanding of the requirements for structural metadata for digital resources, of the role of METS in "wrapping" metadata and digital objects, and of the role of IMS Content Packaging in the exchange of reusable learning resources.

Standards:

Metadata Encoding and Transmission Standard (METS)
<http://www.loc.gov/standards/mets/>
Available 2008-04-28

IMS Content Packaging.
<http://www.imsproject.org/content/packaging/>
Available 2008-04-28

Guidance:

METS, Wikipedia
<http://en.wikipedia.org/wiki/METS>
Available 2008-04-28

Metadata for digital libraries: state of the art and future directions, Richard Gartner, JISC Techwatch report,
<http://www.jisc.ac.uk/whatwedo/services/services_techwatch/techwatch/Techwatch_ic_reports2008_published.aspx>
Available 2008-05-19

6.2.5. Collection-Level Description

A digital resource is created not in isolation but as part of a digital collection, and should be considered within the context of that collection and the development of the collection. Indeed, collections themselves are seen as components around which many different types of digital services might be constructed.

Collections should be described so that a user can discover important characteristics of the collection and so that collections can be integrated into the wider body of existing digital collections and into digital services operating across these collections.

Projects should display awareness of initiatives to enhance the disclosure and discovery of collections, such as programme-, community-, sector- or domain-wide, national, or international inventories of digitisation activities and of digital cultural content. Projects should contribute metadata to such services where appropriate.

Projects should provide collection-level descriptions using an appropriate metadata schema. Projects should display awareness of the Dublin Core Collections Application Profile and work by the MICHAEL project on collection level description.

Standards:

Dublin Core Collection Description Application Profile
<http://dublincore.org/groups/collections/>
Available 2008-04-28

References:

MICHAEL Data Model
<http://www.michael-culture.eu/documents/MICHAELDataModelv1_0.pdf>
Available 2008-04-28

Minerva: Deliverable D3.2: Inventories, discovery of digitised content & multilingual issues: Feasibility survey of the common platform
<http://www.minervaeurope.org/intranet/reports/D3_2.pdf>
Available 2008-04-28

RSLP Collection Description
<http://www.ukoln.ac.uk/metadata/rslp/>
Available 2008-04-28

Guidance:

Minerva: Deliverable D3.1: Inventories, discovery of digitised content & multilingual issues: Report analysing existing content
<http://www.minervaeurope.org/intranet/reports/D3_1.pdf>
Available 2008-04-28

6.2.6. Terminology Standards

Effective transmission of the information conveyed in metadata records requires more than a shared understanding of the metadata schema in use and its constituent metadata elements. It also depends on establishing shared understanding of the terms used as values of those metadata elements, either by the adoption of common terminologies or by adopting different terminologies where the relationships between terms are clearly defined.

Projects **should** use recognised multilingual terminological sources to provide values for metadata elements where possible. Only if no standard terminology is available, local terminologies **may** be considered. Where local terminologies are deployed, information about the terminology and its constituent terms and their meaning **must** be made publicly available.

The use of a terminology in metadata records, either standard or project-specific, **must** be indicated unambiguously in the metadata records.

Collection-level metadata records **should** make use of the terminologies recommended for use with the Minerva collection-level description schema.

Standards:

Minerva: Deliverable D3.2: Inventories, discovery of digitised content & multilingual issues: Feasibility survey of the common platform
<http://www.minervaeurope.org/intranet/reports/D3_2.pdf>
Available 2008-04-28

6.3. Web Ontologies

Projects **may** wish to take advantage of the capacities to share and reuse data on the Web that are provided by a family of specifications coordinated by W3C's Semantic Web activity. RDF provides a standard way of expressing simple descriptions of resources.

Projects **may** wish to make use of Web-based ontologies created using the Web Ontology Language (OWL). OWL builds on RDF and RDF Schema to add a richer vocabulary to describe properties and classes to facilitate the creation of machine-processable definitions of basic concepts and the relationships among them.

Projects **may** wish to explore the potential for semantic interoperability offered by established ontologies such as the CIDOC Conceptual Reference Model (CRM) or the ABC Ontology/Model developed within the Harmony Project.

The CRM provides a common and extensible semantic framework that any cultural heritage information can be mapped to, and can provide a model for mediating between different sources of information.

The ABC Ontology is a top-level ontology to facilitate interoperability between metadata schemas within the digital library domain.

Standards:

W3C Semantic Web Activity
<http://www.w3.org/2001/sw/>
Available 2008-04-28

Web Ontology Language (OWL), W3C,
<http://www.w3.org/2001/sw/WebOnt/>
Available 2008-04-28

OWL Web Ontology Language Reference, W3C, 10 Feb 2004,
<http://www.w3.org/TR/owl-ref/>
Available 2008-04-28

CIDOC Conceptual Reference Model (CRM)
<http://cidoc.ics.forth.gr/>
Available 2008-04-28

Guidance:

RDF Primer
<http://www.w3.org/TR/rdf-primer/>
Available 2008-04-28

OWL Web Ontology Language Overview
<http://www.w3.org/TR/owl-features/>
Available 2008-04-28

The ABC Ontology and Model
<http://metadata.net/harmony/JODI_Final.pdf>
Available 2008-04-28

Semantic Web, Wikipedia
<http://en.wikipedia.org/wiki/Semantic_web>
Available 2008-04-28

Web Ontology Language, Wikipedia
<http://en.wikipedia.org/wiki/Web_Ontology_Language>
Available 2008-04-28

7. Publication

It is expected that end-user access to resources will be primarily through the use of Internet protocols. Preparation for publication requires the processing of the "digital master" to generate digital objects suitable for use in the Internet context, typically by reducing quality in order to generate files of sizes suitable for transfer over networks.

Also, video and audio may be made available either for download or for streaming, which means that instead the entire file being transferred before playback can start, a small buffer space is created on the user's computer, and data is transmitted into the buffer. As soon as the buffer is full, the streaming file starts to play, while more data continues to be transmitted.

Consideration **must** be given to the fact that variations exist in

- the types of hardware device and client software employed by users
- the levels of bandwidth restriction within which users operate

To maximise potential audience reach, projects **should** make resources available in alternative sizes or formats or at alternative resolutions/bit-rates. Project **should** periodically review the criteria on which decisions about delivery formats and parameters are based.

Note: The following recommendations on delivery formats should be read in conjunction with the requirements for file formats for storage of resources (see 5.1).

7.1. Processing for delivery

7.1.1. Delivery of Text

Character Encoding

The character encoding used in text-based documents **should** be transmitted in the HTTP header, and also recorded within documents as appropriate.

Note that some XML-based protocols may mandate the use of a specified character encoding, e.g. the OAI Protocol for Metadata Harvesting requires the use of the UTF-8 character encoding.

Guidance:

Jukka Korpela, A Tutorial on Character Code Issues
<http://www.cs.tut.fi/~jkorpela/chars.html>
Available 2008-04-28

Document Formats

Text-based content **must** be delivered as XHTML 1.0 or HTML 4 (or subsequent versions), though the use SGML or XML formats conforming to other DTDs or Schemas **may** sometimes be appropriate.

In some cases, delivery in proprietary formats such as PDF, ODF, RTF or Microsoft Word **may** be appropriate as a supplementary format to XHTML/HTML, but projects **must** ensure that accessibility issues have been addressed.

It should also be noted that the Open XML (OOXML) and the OpenDocument formats are in the process of being standardised as open standards.

Standards:

HTML 4.01 HyperText Markup Language
<http://www.w3.org/TR/html401/>
Available 2008-04-28

XHTML 1.0 The Extensible HyperText Markup Language
<http://www.w3.org/TR/xhtml1/>
Available 2008-04-28

Information technology -- Open Document Format for Office Applications (OpenDocument) v1.0
<http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43485>
Available 2008-04-28

Portable Document Format (PDF)
<http://www.adobe.com/products/acrobat/adobepdf.html>
Available 2008-04-28

Guidance:

OpenDocument, Wikipedia,
<http://en.wikipedia.org/wiki/OpenDocument>
Available 2008-04-28

Office Open XML, Wikipedia,
<http://en.wikipedia.org/wiki/Ecma_Office_Open_XML>
Available 2008-04-28

PDF, Wikipedia,
<http://en.wikipedia.org/wiki/Pdf>
Available 2008-04-28

7.1.2. Delivery of Still Images

Photographic images

Photographic images **must** be provided on the Web as JPEG/SPIFF format.

Consideration **should** be given to providing various sizes of image to offer readability appropriate to the context of use. IPR issues may also contribute to decisions about the size and quality of image provided.

Thumbnail images **should** be provided at a resolution of 72 dpi, using a bit depth of 24-bit colour or 8-bit greyscale, and using a maximum of 100-200 pixels for the longest dimension (Source: EMII-DCF).

Images for full-screen presentation **should** be provided at a resolution of 150 dpi, using a bit depth of 24-bit colour or 8-bit greyscale and using a maximum of 600 pixels for the longest dimension. This resolution remains lower than that required for high quality print reproduction (Source: EMII-DCF).

Graphic non-vector images

Images **should** be delivered on the Web using Graphical Interchange Format (GIF) or Portable Network Graphics (PNG) format.

Graphic vector images

Images **should** be delivered on the Web using the Scalable Vector Graphics (SVG) formats.

7.1.3. Delivery of Video

Consideration **should** be given to the possibility that users' access to video may be constrained by bandwidth restrictions and it may be appropriate to provide a range of files or streams of different quality.

Downloading

Video for download **should** be delivered on the Web using the MPEG-1 format or the proprietary Microsoft Audio Video Interleave (AVI), Windows Media Video (WMV) or Apple Quicktime formats.

The Reference Website for MPEG
<http://www.mpeg.org/>
Available 2008-04-28

Streaming

Video for streaming **should** be delivered on the Web using Microsoft Advanced Streaming Format (ASF), Windows Media Video (WMV) or Apple Quicktime formats.

7.1.4. Delivery of Audio

Consideration **should** be given to the possibility that users' access to audio may be constrained by bandwidth restrictions and it may be appropriate to provide a range of files or streams of different quality.

Downloading

Audio **should** be delivered on the Web in a compressed form, using the MPEG Layer 3 (MP3) format or the proprietary RealAudio (RA) or Microsoft Windows Media Audio (WMA) formats. A bitrate of 256 Kbps should be used where near CD quality sound is required; a bitrate of 160 Kbps provides good quality.

Audio **may** be delivered in uncompressed forms using the Microsoft WAV/AIFF or Sun AU formats.

Streaming

Audio for streaming **should** be delivered on the Web using the MPEG Layer 3 (MP3) format or the proprietary RealAudio (RA) or Microsoft Windows Media Audio (WMA) formats.

Standards:

Uniform Resource Identifiers (URI)
<http://www.w3.org/Addressing/>
Available 2008-04-28

Guidance:

MP3, Wikipedia,
<http://en.wikipedia.org/wiki/MP3>
Available 2008-04-28

WMV, Wikipedia,
<http://en.wikipedia.org/wiki/Wmv>
Available 2008-04-28

Quicktime, Wikipedia,
<http://en.wikipedia.org/wiki/Quicktime>
Available 2008-04-28

Audio streaming, Wikipedia,
<http://en.wikipedia.org/wiki/Audio_streaming>
Available 2008-04-28

7.1.5. Identification

Digitised resources **should** be unambiguously identified and uniquely addressable directly from a user's Web browser. It is important, for example, that the end user has the capability to directly and reliably cite an individual resource, rather than having to link to the Web site of a whole project. Projects **should** make use of the Uniform Resource Identifier (URI) for this purpose, and **should** ensure that the URI is reasonably persistent. Such URIs **should not** embed information about file format, server technology, organisational structure of the provider service or any other information that is likely to change within the lifetime of the resource.

Where appropriate, projects **may** wish to consider the use of OpenURLs, Digital Object Identifiers or of persistent identifiers based on another identifier scheme.

Projects **may** also wish to ensure that logical sets within the resources they are providing are uniquely and persistently addressable.

Standards:

Uniform Resource Identifiers (URI)
<http://www.w3.org/Addressing/>
Available 2008-04-28

ANSI/NISO Z39.88 -2004, The OpenURL Framework for Context-Sensitive Services
<http://www.niso.org/standards/standard_detail.cfm?std_id=783>
Available 2008-04-28

Digital Object Identifier (DOI)
<http://www.doi.org/>
Available 2008-04-28

Guidance:

Digital object identifier, Wikipedia,
<http://en.wikipedia.org/wiki/Digital_Object_Identifier>
Available 2008-04-28

OpenURL, Wikipedia,
<http://en.wikipedia.org/wiki/OpenURL>
Available 2008-04-28

7.2. 3D and Virtual Reality Issues

Projects making use of three-dimensional virtual reality (VR) 'fly throughs' and models **must** consider the needs of users accessing their site using typical computers and modem connections.

These models are typically used in the reconstruction of buildings and other structures or in simulating whole areas of a landscape. Traditionally, models have been constructed and displayed using powerful computer workstations, and this continues to be the case for the most detailed. For projects that are required to deliver the results of their work to a large audience via the Internet, such highly detailed models may be unhelpful. Nevertheless, there is scope for usefully incorporating less complex models into the Web sites made available to users.

In generating these models, projects **must** be aware that the majority of their users for the foreseeable future will continue to access the Internet using a 56k modem or a shared connection, rather than any higher bandwidth technology. Similarly, the specifications of the computers being used by typical visitors are likely to be significantly lower than those of the machines on which projects generate and test any such models. Projects **must** therefore consider the usability of their models in such conditions, and **must** test them using typical modem connections and home, school, or library computer systems with a variety of typical operating systems and browsers.

Standards in this area continue to evolve, but projects **should** produce VR models compatible with the X3D specification.

Apple's QuickTime VR (QTVR) is not a true 3D image format, but does offer some useful functionality. Projects which do not require the full functionality of X3D **may** wish to consider using QTVR instead.

Standards:

Web3D Consortium
<http://www.web3d.org/>
Available 2008-04-28

X3D
<http://www.web3d.org/x3d/>
Available 2008-04-28

QuickTime VR
<http://www.apple.com/quicktime/technologies/qtvr/>
Available 2008-04-28

7.3. Geographic Information Systems

Much cultural content has a grounding in place, and this offers one powerful means by which content might be grouped or retrieved. Geographic Information Systems (GIS) are software applications specifically designed to store, manipulate and retrieve place-based information, and they are increasingly widely deployed within the cultural heritage sector.

It is not necessary, however, for every project that wishes to store place-based information, or to include a location map on their web site, to install and maintain a GIS. Place-based information may be stored (although not necessarily fully manipulated or re-used) within a traditional database, and simple images of location maps etc may be created by various means. Projects **must** ensure that they can support a GIS implementation in the future, even if this is not planned within the project. The OAI-PMH harvesting of metadata (using a schema such as DC.Culture) can then allow the presentation of data through an external GIS system (see Section 8.1).

For those projects that do require rich interaction with place-based information, such as that potentially offered by a GIS, the following **must** be borne in mind:

Projects seeking to employ a GIS **must** obtain appropriate permissions for use of any map data from third parties, ensuring that licences extend to delivering services to their defined audiences via their selected delivery channels.

Projects **must** ensure that data sets combined for the purposes of delivering their service are of similar scale and resolution, and appropriate for being used together in this manner.

Commercial GIS products selected for use **should** comply with emerging industry standards from the Open GIS Consortium.

Projects **must** make use of and declare use of an appropriate standard co-ordinate reference system when recording spatial data.

Projects **must** make use of and declare use of appropriate national standards for the recording of street addresses.

Standards:

OpenGIS Consortium
<http://www.opengis.org/>
Available 2008-04-28

Guidance:

Archaeology Data Service GIS Guide to Good Practice
<http://ads.ahds.ac.uk/project/goodguides/gis/>
Available 2008-04-28

7.4. Web Sites

Project resources **must** be accessible using a Web browser. This will normally be achieved using HTML or XHTML and the HTTP 1.1 protocol (although many other file formats can be delivered over HTTP). If other protocols are used the information **must** be available to provide access by a Web browser.

Projects **should** seek to provide maximum availability of their project Web site. Significant periods of unavailability **should** be accounted for to the funding programme.

Standards:

Hypertext Transfer Protocol, HTTP/1.1, W3C,
<http://www.w3.org/Protocols/HTTP/>
Available 2008-04-28

Guidance:

Hypertext Transfer Protocol, Wikipedia,
<http://en.wikipedia.org/wiki/Http>
Available 2008-04-28

7.4.1. Accessibility

Projects **must** be accessible by a variety of browsers, hardware systems, automated programs and end-users.

Web sites **must** be accessible to a wide range of browsers, mobile devices, etc. Web sites **must** be usable by browsers that support W3C recommendations such as HTML/XHTML, Cascading Style Sheets (CSS) and the Document Object Model (DOM). Projects that make use of proprietary file formats and browser plug-in technologies **must** ensure that their content is still usable on browsers that do not have the plug-ins. As a result, the use of technologies such as JavaScript and Macromedia Flash in navigation of the site must be carefully considered.

The appearance of a Web site **should** be controlled by use of style sheets in line with W3C architecture and accessibility recommendations. The latest version of Cascading Style Sheets (CSS) recommended by W3C (currently CSS 2) should be used, although, due to incomplete support by browsers, not all features defined in CSS 2 may be usable.

Projects **must** implement W3C Web Accessibility Initiative (WAI) recommendations and so ensure a high degree of accessibility for people with disabilities. In is normally recommended that projects **must** achieve WAI WCAG 1.0 level A conformance; projects **should** aim to achieve WCAG 1.0 level AA conformance. However it should be noted that WCAG 2.0 is currently being developed, so projects **should** bear in mind the implications of migrating to compliance with this new set of guidelines.

Standards:

Cascading Style Sheets (CSS), Level 2, W3C,
<http://www.w3.org/TR/REC-CSS2/>
Available 2008-04-28

Web Content Accessibility Guidelines (WCAG) 1.0, W3C,
<http://www.w3.org/TR/WCAG10/>
Available 2008-04-28

Web Content Accessibility Guidelines (WCAG) 2.0, W3C,
<http://www.w3.org/TR/WCAG20/>
Available 2008-04-28

Guidance:

Web Accessibility Initiative (WAI), W3C,
<http://www.w3.org/WAI/>
Available 2008-04-28

Accessible Web Design, RNIB,
<http://www.rnib.org.uk/digital/hints.htm>
Available 2008-04-28

WCAG, Wikipedia,
<http://en.wikipedia.org/wiki/WCAG>
Available 2008-04-28

7.4.2. Security

The machines used to deliver projects **must** be operated in as secure a manner as possible. The advice in operating system manuals concerning security **must** be followed. All known security patches **must** be applied.

Machines **should** be configured to run only the minimum number of network services. Machines **should** be placed behind a firewall if possible, with access to the Internet only on those ports that are required for the project being delivered.

Projects **should** demonstrate awareness of the codes of practice provided by ISO/IEC 17799:2000. The management and use of any personal information **must** conform to relevant national legislation.

Where sensitive information is being passed from a client to a server across the network, projects **must** use Secure Sockets Layer (SSL) to encrypt the data. This includes the transfer of usernames

and passwords, credit card details and other personal information. Note that the use of SSL also provides the end-user with an increased level of confidence in the authenticity of the service.

Standards:

Secure Sockets Layer (SSL) 3.0
<http://wp.netscape.com/eng/ssl3/>
Available 2008-04-28

Guidance:

Transport Layer Security, Wikipedia,
<http://en.wikipedia.org/wiki/Transport_Layer_Security>
Available 2008-04-28

7.4.3. Authenticity

Project specific domain names **should** be registered in the Domain Name System (DNS). The domain name forms part of the project 'branding' and will help end-users identify the authenticity of the content being delivered. Domain names **should** therefore be clearly branded with either the name of the project or the organisation delivering the project.

In some situations it may be appropriate to secure the network connection between the client and the server using Secure Sockets Layer (SSL) to give end-users increased confidence that they are exchanging information with the correct project Web site.

Guidance:

Domain Names System (DNS) Resources Directory
<http://www.dns.net/dnsrd/>
Available 2008-04-28

7.4.4. User Authentication

Some projects **may** wish to limit access to parts of their resources (for example to very high-resolution images or maps, etc.) to authenticated users only. User authentication is an important tool for ensuring that only legitimate users can access the project's online resources.

If projects choose to implement user authentication for selected materials it **should** be based on a username and password combination. In the case of Web-based projects, HTTP Basic Authentication **must** be used to pass the username/password combination from the browser to the server.

In some cases IP-based authentication (comparing the IP address of the client against a list of known IP addresses) **may** be an appropriate alternative to usernames and passwords. However, the use of this authentication method is strongly discouraged since the growth in the use of dynamic IP addressing by many Internet Service Providers will make it very difficult to manage a list of approved IP addresses. In addition support for mobile users and users behind firewalls will also make IP authentication difficult to manage.

Projects **may** choose to make use of third party authentication services to manage usernames and passwords on their behalf, if appropriate.

Standards:

Hypertext Transfer Protocol, HTTP/1.1, W3C,
<http://www.w3.org/Protocols/HTTP/>
Available 2008-04-28

8. Use of Resources

The collections developed by a digitisation project from part of a larger corpus of material. To support the discovery of resources within that corpus, for each collection, projects **must** consider exposing metadata about their resources so that it can be used by other applications and services, using one or more of the protocols or interfaces described in the following sub-sections.

The precise requirements in terms of what metadata should be provided and how that metadata should be exposed will depend on the nature of the resources created and the applications and services with which that metadata is shared.

Projects **should** expose one or more collection-level metadata records describing their collections as units. Projects **may** expose item-level metadata records describing individual digital resources within their collection(s).

Both collection-level and item-level metadata records **should** include a statement of the conditions and terms of use of the resource.

In order to facilitate potential exchange and interoperability between services, projects **should** be able to provide item level descriptions in the form of simple, unqualified Dublin Core metadata records and **may** provide item-level descriptions conforming to the experimental DC.Culture schema.

Where items are "learning resources" or resources of value to the learning and teaching communities, projects **should** also consider providing descriptions in the form of IEEE Learning Object Metadata.

Projects **should** also display awareness of any additional requirements to provide metadata imposed by their operating context (e.g. national government metadata standards).

Projects **should** maintain awareness of any rights issues affecting their metadata records.

Standards:

Dublin Core Metadata Element Set, Version 1.1
<http://dublincore.org/documents/dces/>
Available 2008-04-28

DC.Culture
<http://www.minervaeurope.org/DC.Culture.htm>
Available 2008-04-28

IEEE Learning Object Metadata
<http://ltsc.ieee.org/wg12/>
Available 2008-04-28

Guidance:

Dublin Core, Wikipedia,
<http://en.wikipedia.org/wiki/Dublin_core>
Available 2008-04-28

Learning object metadata, Wikipedia,
<http://en.wikipedia.org/wiki/Learning_object_metadata>
Available 2008-04-28

8.1. Metadata Harvesting

Projects **should** demonstrate awareness of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) as a means of making their metadata available to service providers.

Projects **may** consider making their metadata available for harvesting by setting up OAI compliant metadata repositories. Projects that do establish such repositories **should** consider inclusion of a statement of the rights held in their metadata to ensure they retain ownership rights in their metadata.

Standards:

Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
<http://www.openarchives.org/>
Available 2008-04-28

Guidance:

OAI FAQs
<http://www.openarchives.org/documents/FAQ.html>
Available 2008-04-28

Open Archives Initiative, Wikipedia,
<http://en.wikipedia.org/wiki/Open_Archives_Initiative>
Available 2008-04-28

8.2. Distributed Searching

Projects **may** need to demonstrate awareness of the Search/Retrieve Web Service (SRW/SRU) protocol, which builds on Z39.50 semantics to deliver similar functionality using Web Service technologies.

Projects **may** wish to consider the potential of Z39.50, a network protocol that allows searching of heterogeneous databases and retrieval of data. Z39.50 is most often used for retrieving bibliographic records, although there are also some non-bibliographic implementations. Projects that do use Z39.50 **must** display awareness of the Bath Profile and its relevance to cross-domain interoperability.

Standards:

SRW: Search/Retrieve Web Service
<http://www.loc.gov/standards/sru/>
Available 2008-04-28

Z39.50 Maintenance Agency
<http://www.loc.gov/z3950/agency/>
Available 2008-04-28

Bath Profile
<http://www.collectionscanada.ca/bath/tp-bath2-e.htm>
Available 2008-04-28

Guidance:

Z39.50, Wikipedia,
<http://en.wikipedia.org/wiki/Z39.50>
Available 2008-04-28

What is SRW/U?, Techessence,
<http://techessence.info/node/48>
Available 2008-04-28

8.3. Syndication / Alerting

Projects **should** demonstrate awareness of the RSS family of specifications, including the related Atom format. RSS provides a mechanism for sharing descriptive metadata, typically in the form of a list of items, each containing a brief textual description along with a link to the originating source for expansion. Although originally developed as a mechanism for news alerts, RSS has now become a well-established mechanism for syndicating content.

RSS 2.0 can be used for syndicating audio and video content to various devices, including portable media players, such as iPods. The term ‘podcasting’ or ‘vodcasting’ has been used to describe this type of syndication.

Standards:

RDF Site Summary (RSS) 1.0
<http://purl.org/rss/1.0/spec>
Available 2008-04-28

RSS 2.0
<http://blogs.law.harvard.edu/tech/rss>
Available 2008-04-28

The Atom Syndication Format
<http://atompub.org/2005/04/18/draft-ietf-atompub-format-08.html>
Available 2008-04-28

Guidance:

RSS, Wikipedia
<http://en.wikipedia.org/wiki/Rss>
Available 2008-04-28

Podcast, Wikipedia
<http://en.wikipedia.org/wiki/Podcast>
Available 2008-04-28

Video podcast, Wikipedia
<http://en.wikipedia.org/wiki/Vodcast>
Available 2008-04-28

8.4. Web Services

Projects **should** demonstrate awareness of the Web Services family of specifications, especially SOAP version 1.2 and the Web Services Description Language (WSDL).

For network services not covered by the specific protocols discussed above, consideration **should** be given to the use of SOAP, though use of the REST architectural style through HTTP 1.1 GET or POST requests to return XML documents **may** be appropriate.

Projects **may** also be required to show awareness of the Universal Description, Discovery & Integration (UDDI) specification.

Standards:

SOAP Version 1.2 Part 1: Messaging Framework
<http://www.w3.org/TR/soap12-part1/>
Available 2008-04-28

Web Services Description Language (WSDL) 1.1
<http://www.w3.org/TR/wsdl>
Available 2008-04-28

Hypertext Transfer Protocol, HTTP/1.1
<http://www.w3.org/Protocols/HTTP/>
Available 2008-04-28

Guidance:

SOAP Version 1.2 Part 0: Primer
<http://www.w3.org/TR/soap12-part0/>
Available 2008-04-28

Web service, Wikipedia
<http://en.wikipedia.org/wiki/Web_services>
Available 2008-04-28

Universal Description Discovery and Integration, Wikipedia
<http://en.wikipedia.org/wiki/UDDI>
Available 2008-04-28

9. Reuse and Re-purposing

Other parties may want to repackage and re-purpose material that has been developed by digitisation projects. This could include end users who wish to reuse the resources in their own environment and to reflect their own personal preferences and third party organisations who may wish to provide access to resources ito new target audiences or provide value-added services. In order to facilitate this re-use the implementation of standards will be important.

9.1. Learning Resource Creation

Projects **should** consider the potential re-use of the resources they create, and recognise that end users or third parties may wish to extract elements of a given resource and repackage them with parts of other resources from their own collections and from other sources.

An important area in which this is likely to happen is the educational sector. In the global educational community, a number of initiatives are underway to create tools for managing educational resources. Some of this effort is concentrating upon the description of content such as that created by digitisation programmes.

Projects that develop learning resources **must** demonstrate awareness of the IEEE Learning Object Metadata (LOM) standard and **should** consider providing LOM descriptions of their learning resources.

Project **should** track the work of the IMS consortium in developing specifications to support interoperability amongst learning technology systems. Projects that develop learning resources **should** consider the use of IMS Content Packaging to facilitate access to those resources by users of Virtual Learning Environment systems.

Standards:

IEEE Learning Object Metadata
<http://ltsc.ieee.org/wg12/>
Available 2008-04-28

IMS Global Learning Consortium, Inc.
<http://www.imsproject.org/>
Available 2008-04-28

IMS Content Packaging.
<http://www.imsproject.org/content/packaging/>
Available 2007-10-30

Guidance:

Learning object metadata, Wikipedia
<http://en.wikipedia.org/wiki/Learning_object_metadata>
Available 2008-04-28

9.2. Reuse By Third Party Services

Use of open standards can make it easier for third parties to re-use materials. However legitimate use will be reliant on corresponding terms and conditions, such as Creative Commons licences.

Standards:

Creative Commons,
<http://www.creativecommons.org/>
Available 2008-04-28

Guidance:

Creative Commons, Wikipedia,
<http://en.wikipedia.org/wiki/Creative_Commons>
Available 2008-04-28

10. Intellectual Property Rights and Copyright

Projects **must** respect intellectual property rights held in the materials they work with, including:

- the rights of the owners of the source materials that are digitised;
- the rights of the owners of the digital resources;
- the rights or permissions granted to a service provider to make the digital resources available;
- the rights or permissions granted to the users of the digital resources.

Projects **must** also respect any rights arising from the particular terms and conditions of any digitisation programme within which they are operating.

Care is particularly advisable in the circumstances below:

- **Published material.** Publishers are unlikely to give permission to digitise in-copyright material unless this is of some advantage to them. Older material may be out of copyright but the project is responsible for confirming this.
- **In-house productions.** The rights in any work undertaken by an institution's staff as part of their normal duties remains the property of that institution. In some academic institutions these rights may not have been asserted, and authors may have assigned them to external publishers. Unpaid volunteers retain the copyright of their work unless they sign away their rights.
- **Institutions commissioning work.** This work, for example photography, will normally have secured reproduction rights, but this may not have extended to digitisation unless specifically stated in the agreement. Projects will only have copyright on digitised material if this permission is secured.
- **Gifts, bequests and loans.** These may have particular conditions attached to them that affect their availability for digitisation.

10.1. Identifying, Recording and Managing Intellectual Property Rights

In order to manage rights held in cultural resources, projects **must** first identify and record what rights exist in the materials.

Where necessary projects **must** negotiate with rights holders to obtain permission to use materials.

Projects **must** record the permissions granted in licences, which specify the nature and scope of the content, the ways in which it can be used, the geographical extent of the rights, the duration of the licence and, where appropriate, a fee.

Projects **must** monitor licensing arrangements and ensure that licences are re-negotiated as required.

Guidance:

Creating Digital Resources for the Visual Arts: Standards and Good Practice
<http://vads.ahds.ac.uk/guides/creating_guide/contents.html>
Available 2008-04-28

Intellectual property Rights Overview
<http://www.jisclegal.ac.uk/ipr/IntellectualProperty.htm>
Available 2008-04-28

UK Intellectual Property Office
<http://www.ipo.gov.uk/home.htm>
Available 2008-04-28

World Intellectual Property Organization
<http://www.wipo.org/>
Available 2008-04-28

10.2. Safeguarding Intellectual Property Rights

Having identified property rights and negotiated licences, projects **must** ensure that their rights and the rights of other parties are protected, by taking steps to ensure that there is no unauthorised use of materials.

In the network environment, every transaction that involves intellectual property is by its nature a rights transaction. The expression of these 'Terms of Availability' or 'Business Rules' is dependent on 'rights metadata' – data which identifies unambiguously and securely the intellectual property itself, the specific rights which are being granted (for example to read, to print, to copy, to modify) and the users or potential users.

Projects **should** maintain data about the rights that they hold and acquire in an internally consistent form, so that they can be shared in a standard format.

The type of information required includes:

- The identification of the resource itself.
- The name of the person or organisation granting the rights.
- The precise right or rights that are being granted (including, for example, whether modification is permitted) – and any specific exclusions.
- The period of time for which rights are granted.
- The user group or groups permitted to use the resource.
- Any obligations (including but not limited to financial obligations) that users of the resource may incur.

10.2.1. Creative Commons

The Creative Commons initiative has released of a set of copyright licenses that are free for public use, and enable people to share their works and either to dedicate their creative works to the public domain or to retain their copyright while licensing them as free for certain uses, on certain conditions.

Projects **may** wish to assign a Creative Commons licence to their resources.

Standards:

Creative Commons,
<http://www.creativecommons.org/>
Available 2008-04-28

Guidance:

Creative Commons, Wikipedia,
<http://en.wikipedia.org/wiki/Creative_Commons>
Available 2008-04-28

10.2.2. E-Commerce

It is common for public sector content creation programmes to specify that content created must be made available free to users at the point of access, at least for educational purposes. In some cases programmes also encourage or require projects to generate revenue from the materials created.

Projects **must** follow programme requirements regarding access to and use of resources created.

Projects **must** ensure that adequate protection is given to all intellectual property rights.

10.2.3. Watermarking and Fingerprinting

Projects **should** give consideration to watermarking and fingerprinting the digital material they produce.

Watermarking is the embedding of a permanent mark within a file that can subsequently be used to prove image origination or image copyright. This is normally achieved by integrating the watermark with the image data in such a way that it is virtually impossible to remove. Watermarks can be visible, invisible or a combination of both. In all cases the watermark is introduced in such a way that there is minimum distortion of the original image. Invisible watermarks **must** be able to withstand the image being cropped, rotated, compressed or transformed.

As well as watermarking images before they are distributed, images can be fingerprinted dynamically at delivery time i.e. as the image is downloaded from a Web site. When this is done, other information such as username, date, time, IP address etc. can be encoded as part of the watermark. This makes each instance of download unique and traceable through a transaction database enabling tracking of who is downloading images. Similar techniques can be used in audio and video media.

Guidance:

Purloining and Pilfering, Web Developers Virtual Library
<http://www.wdvl.com/Authoring/Graphics/Theft/>
Available 2008-04-28

Appendix 1 Acknowledgements

This document is based primarily on four sources.

- the *NOF-digitise Technical Standards and Guidelines* (Version 5, February 2003), that were developed on behalf of the UK New Opportunities Fund (NOF), by UKOLN, University of Bath, in association with Resource: The Council for Museums, Archives & Libraries (now known as MLA). <http://www.peoplesnetwork.gov.uk/content/technical.asp>.
- additional information provided to NOF-digitise projects in support of the Standards and Guidelines by the NOF-digitise Technical Advisory Service, operated for NOF by UKOLN and the Arts and Humanities Data Service (AHDS), in the form of the programme manual, briefing papers and FAQs. <http://www.ukoln.ac.uk/nof/support/manual/> and <http://www.ukoln.ac.uk/nof/support/help/faqs/>.
- the *Framework Report* (September 2003), published by the European Museums' Information Institute – Distributed Content Framework (EMII-DCF) project, particularly the Data Capture Model in Chapter 16. <http://www.emii-dcf.org/dokument/frame.pdf>.
- the *Good Practice Handbook* (Version 1.2, November 2003), developed by the Minerva project (Working Group 6). http://www.minervaeurope.org/structure/workinggroups/goodpract/document/bestpracticehandbookv1_2.pdf.
- DCC Diffuse standards. <http://www.dcc.ac.uk/diffuse/>.

It also draws on a number of other sources:

The Institute of Museum and Library Services' Framework of Guidance for Building Good Digital Collections
<http://www.imls.gov/pubs/forumframework.htm>

The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials, <http://www.ninch.org/programs/practice/>

Research Libraries Group Cultural Materials Initiative: Recommendations for Digitizing for RLG Cultural Materials, <http://www.rlg.ac.uk/culturalres/prospective.html>

Research Libraries Group Cultural Materials Initiative: Description Guidelines
<http://www.rlg.ac.uk/culturalres/descguide.html>

Canadian Heritage Standards and Guidelines for Digitization Projects
<http://www.pch.gc.ca/progs/pcce-ccop/pubs/ccop-pcceguide_e.pdf>

Working with the Distributed National Electronic Resource (DNER): Standards and Guidelines to Build a National Resource ,
<http://www.jisc.ac.uk/index.cfm?name=projman_standards>

JISC Information Environment Architecture Standards Framework
<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/standards/>

The Public Libraries Managing Advanced Networks (PULMAN) Guidelines
<http://www.pulmanweb.org/DGMs/DGMs.htm>

Version 1.0 of this document was edited by Pete Johnston, UKOLN, with contributions from the MINERVA WP4 Working Group: (Eelco Bruinsma, Consultant, NL; Rob Davies, MDR Partners / PULMAN Project, UK; David Dawson, MLA, UK; Bert Degenhard Drenth, Adlib Information Systems / EMII-DCF Project, NL; Giuliana De Francesco, Ministero per i beni e le attività culturali, IT; Muriel Foulonneau, Relais Culture Europe / EMII-DCF Project, FR; Gordon McKenna, mda / EMII-DCF project, UK; Paul Miller, UKOLN, UK; Maureen Potter, ERPANET Project, NL; Jos Taekema, Digital Erfgoed Nederland, NL and Chris Turner, MLA, UK).

This version was edited by Brian Kelly, UKOLN based on input received from members of the Minerva project.

Appendix 2 About This Document

This document has sought to provide a core set of guidelines, rather than to attempt to reflect the different requirements of many different programmes and projects. The implementers of digitisation programmes and projects will need to adapt these guidelines to the specific contexts in which they are operating, to select, to customise and to supplement as required. However, it is hoped that as a core, they can provide a starting point that is useful in many different contexts.

Maintenance

These Guidelines were developed by the MINERVA project. All comments and suggestions for changes and updates should be submitted to the MINERVA Project.

Guidance:

MINERVA Project Website
<http://www.minervaeurope.org/>

Links to Resources

The resources cited in this document have been bookmarked in the del.icio.us bookmarking service at the URL <http://del.icio.us/lisbk/minerva-2008>. This online resource will be used to ensure that links are still operational during the lifetime of this document.

Links to Guidance in Wikipedia

A number of resources of further guidance about the standards described in this document link to information provided in Wikipedia. As with all information resources, care is needed when interpreting information provided on the Internet. However since many digital library professionals are responsible for creating and maintaining information provided in Wikipedia related to digital library standards, it is felt that Wikipedia will often provide a more up-to-date source of information than reports which have been published and which are no longer being updated.

Appendix 3 Exploiting the Potential of Web 2.0

The popularity of many Web 2.0 services and the popularity of social networking services is generating much interest in the ways in which Web 2.0 can be used to enhance the quality and take-up of networked services within the cultural heritage sector.

Many Web 2.0 services are based on use of lightweight standards (such as RSS and Atom) and of an agile development environment. In addition commercially hosted Web 2.0 services (such as the portfolio of services provided by well-established companies such as Google and Yahoo!, can be used to deliver or augment services which may traditionally have had to be developed or hosted in-house.

Many of these services are also demonstrating alternative approaches to Web accessibility to the recommendations described in WAI's WCAG 1.0 guidelines. We are seeing, for example how JavaScript and Ajax technologies can be used to provide more usable and accessible technologies than would be possible if such technologies were not used.

Possible disadvantages of using such services include:

- Potential security and legal concerns e.g. copyright, data protection, etc.
- Potential for data loss or misuse.
- Reliance on third parties with whom there may be no contractual agreements.

However there are risks associated with use of a Web 2.0 approach to the development and deployment of Web 2.0. A number of such risks associated are given below, together with an approach to managing the dangers of such risks.

Risk	**Assessment**	**Management**
Loss of service	Implications if service becomes unavailable. Likelihood of service unavailability.	Use for non-mission critical services. Have alternatives available. Use trusted services. Investigate services.
Data loss	Likelihood of data loss. Lack of export capabilities.	Evaluation of service. Non-critical use. Testing of export.
Performance problems	Slow performance	Testing. Non-critical use.
Lack of inter-operability	Likelihood of application lock-in. Loss of integration & reuse of data.	Evaluation of integration and export capabilities.
Format changes	New formats may not be stable.	Plan for migration or use on a small-scale.
Legal issues	User-generated content may be illegal, breach copyright, etc.	Deploy either approval processes or just-in-time moderation.
User issues	User views on services.	Gain feedback.

Table 1: Risk Assessment And Risk Management Approaches For Use of Web 2.0

Note that in addition to risk assessment of Web 2.0 services, there is also a need to assess the risks of failing to provide such services, such as the missed opportunity costs, the costs of developing services in-house, the risks of losing staff, etc.

Appendix 4 Glossary

A glossary of abbreviations and acronyms used in this document is given below.

AHDS	Arts and Humanities Data Service.
AVI	Audio Video Interleave.
dpi	Dots per inch.
DCC	Digital Curation Centre.
DTD	Document Type Definition.
DWG	Drawing.
EMII	European Museums' Information Institute.
FIAF	International Federation of Film Archives.
GIF	Graphical Interchange Format.
HTML	Hypertext markup Language
IDABC	Interoperable Delivery of European eGovernment Services to public Administrations, Businesses and Citizens.
IETF	Internet Engineering Task Force.
IMS	IMS Global Learning Consortium.
JISC	Joint Information Systems Committee.
JPEG	Joint Photographic Expert Group.
JPEG/SPIFF	JPEG Still Picture Interchange File Format.
METS	Metadata Encoding and Transmission Standard.
MPEG	Moving Picture Experts Group.
NISO	National Information Standards Organization.
OAIS	Open Archival Information System.
OCLC	Online Computer Library Center.
PDF	Portable Document Format.
PLANETS	Planets - Preservation and Long-term Access through NETworked Services
PNG	Portable Network Graphics.
PRINCE 2	PRojects IN Controlled Environments – a project management methodology.
RFC	Request For Comments.
RfP	Request for Proposal.
RLG	Research Libraries Group.
SGML	Standard Generalized Markup Language.
SVG	Scalable Vector Graphics.
SWF	Shockwave Flash.
TASI	Technical Advisory Service for Images.
TEI	text Encoding Initiative.
TIFF	Tagged Image File Format.
TCP/IP	Transmission Control Protocol/Internet Protocol.
XHTML	Extensible HyperText Markup Language.
XML	Extensible Markup Language.