



---

# Technical Guidelines for Digital Cultural Content Creation Programmes

*Working Draft Version 0.06: Revised 25 November 2003*

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

## 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.
- 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.
- 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.
- the *Good Practice Handbook* (Version 1.2, November 2003), developed by the Minerva project (Working Group 6).

NOF-digitise Technical Standards and Guidelines

<<http://www.peoplesnetwork.gov.uk/content/technical.asp>>

NOF-digi Technical Advisory Service Programme Manual

<<http://www.ukoln.ac.uk/nof/support/manual/>>

NOF-digi Technical Advisory Service FAQ

<<http://www.ukoln.ac.uk/nof/support/help/faqs/>>

EMII-DCF Framework Report

<<http://www.emii-dcf.org/dokument/frame.pdf>>

Minerva Working Group 6: Good Practice Handbook

<[http://www.minervaeurope.org/structure/workinggroups/goodpract/document/bestpracticehandbookv1\\_2.pdf](http://www.minervaeurope.org/structure/workinggroups/goodpract/document/bestpracticehandbookv1_2.pdf)>

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](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](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>>

UKOLN is funded by Resource: The Council for Museums, Archives & Libraries, 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

<b>1. Introduction.....</b>	<b>7</b>
1.1. <i>The Purpose of this Document.....</i>	7
1.2. <i>The Role of Technical Standards.....</i>	8
1.3. <i>The Benefits of Deploying Standards.....</i>	9
1.4. <i>The Life Cycle Approach.....</i>	10
1.5. <i>Requirement Levels.....</i>	10
<b>2. Preparation for digitisation.....</b>	<b>12</b>
2.1. <i>Hardware.....</i>	12
2.2. <i>Software.....</i>	12
2.3. <i>Environment.....</i>	12
<b>3. Handling of originals.....</b>	<b>13</b>
3.1. <i>Appropriate movement and manipulation of original material.....</i>	13
3.2. <i>Staff training.....</i>	13
<b>4. The digitisation process.....</b>	<b>14</b>
<b>5. Storage and management of the digital master material.....</b>	<b>15</b>
5.1. <i>File formats.....</i>	15
5.1.1. <i>Text Capture and Storage.....</i>	16
5.1.1.1. <i>Character Encoding.....</i>	16
5.1.1.2. <i>Document Formats.....</i>	16
5.1.2. <i>Still Image Capture and Storage.....</i>	17
5.1.2.1. <i>Raster Images.....</i>	18
5.1.2.2. <i>Vector images.....</i>	19
5.1.3. <i>Video Capture and Storage.....</i>	19
5.1.4. <i>Audio Capture and Storage.....</i>	20
5.2. <i>Media choices.....</i>	20
5.3. <i>Migration strategies.....</i>	21
<b>6. Metadata creation/capture.....</b>	<b>22</b>
6.1. <i>The scope of the metadata.....</i>	22
6.2. <i>Appropriate standards.....</i>	22
6.2.1. <i>Descriptive Metadata.....</i>	23
6.2.2. <i>Administrative Metadata.....</i>	24
6.2.3. <i>Preservation Metadata.....</i>	25
6.2.4. <i>Structural Metadata.....</i>	25
6.2.5. <i>Collection-Level Description.....</i>	25
6.2.6. <i>Terminology Standards.....</i>	26
<b>7. Publication.....</b>	<b>28</b>
7.1. <i>Processing for delivery.....</i>	28
7.1.1. <i>Delivery of Text.....</i>	28
7.1.1.1. <i>Character Encoding.....</i>	28

7.1.1.2.	Document Formats.....	28
7.1.2.	Delivery of Still Images.....	29
7.1.2.1.	Photographic images.....	29
7.1.2.2.	Graphic non-vector images.....	29
7.1.2.3.	Graphic vector images.....	29
7.1.3.	Delivery of Video.....	29
7.1.3.1.	Downloading.....	30
7.1.3.2.	Streaming.....	30
7.1.4.	Delivery of Audio.....	30
7.1.4.1.	Downloading.....	30
7.1.4.2.	Streaming.....	30
7.1.5.	Identification.....	30
7.2.	<i>3D and Virtual Reality Issues</i> .....	31
7.3.	<i>Geographic Information Systems</i> .....	31
7.4.	<i>Web Sites</i> .....	32
7.4.1.	Accessibility.....	32
7.4.2.	Security.....	33
7.4.3.	Authenticity.....	34
7.4.4.	User Authentication.....	34
7.4.5.	Performance Indicators.....	35
<b>8.</b>	<b>Disclosure of resources</b> .....	<b>36</b>
8.1.	<i>Metadata Harvesting</i> .....	36
8.2.	<i>Distributed Searching</i> .....	37
8.3.	<i>Alerting</i> .....	37
<b>9.</b>	<b>Re-use and re-purposing</b> .....	<b>39</b>
9.1.	<i>Learning Resource Creation</i> .....	39
<b>10.</b>	<b>IPR and Copyright</b> .....	<b>40</b>
10.1.	<i>Establishing Copyright</i> .....	40
10.2.	<i>Safeguarding Copyright</i> .....	41
10.2.1.	E-Commerce.....	41
10.2.2.	Watermarking and Fingerprinting.....	41
<b>11.</b>	<b>Summary</b> .....	Errore. Il segnalibro non è definito.



# 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).

A further consideration is the “open-ness” of a standard. This can refer to a number of characteristics of a standard, and the EMII-DCF *Framework Report* 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 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.

## 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 within Work Package 6 of 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 Capture
6. Publication
7. Disclosure
8. Reuse and Re-purposing
9. 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:

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES  
WORKING DRAFT VERSION 0.06  
REVISED 23 DECEMBER 2003

- **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/>>

## 2. Preparation for digitisation

Projects **must** develop a good knowledge of the collections to be digitised and the uses to be made of the digital resources created. 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:*

TASI: Advice: Creating Digital Images

<<http://www.tasi.ac.uk/advice/creating/>>

Available 2003-10-27

The Digitisation Process

<<http://www.ukoln.ac.uk/nof/support/help/papers/digitisation.htm>>

Available 2003-10-27

### 2.1. 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.

### 2.2. 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.

### 2.3. 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.

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES  
WORKING DRAFT VERSION 0.06  
REVISED 23 DECEMBER 2003

## 3. Handling of originals

### 3.1. 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.

*Guidance:*

Joint NPO and RLG Preservation Conference Guidelines for Digital Imaging 28 - 30 September 1998

<http://www.rlg.org/preserv/joint/>

Available 2003-10-27

### 3.2. Staff training

Projects **must** ensure that all staff receive proper training in the use of digitisation hardware and software and in the appropriate handling of physical materials. This ensures that the process is efficient and that any risk to the originals is minimised.

*Guidance:*

TASI: Advice: Creating Digital Images

<http://www.tasi.ac.uk/advice/creating/>

Available 2003-10-27

## 4. 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 three issues:

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

The JISC Image Digitisation Initiative (JIDI), the Arts and Humanities Data Service (AHDS) and the Technical Advisory Service for Images (TASI) all provide further guidance on this topic.

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, the Digital Library Federation, the Council on Library and Information Resources and the Research Libraries Group have recently published some useful Guides to Quality in Visual Resource Imaging.

### **Guidance:**

#### JIDI

<<http://www.ilrt.bris.ac.uk/jidi/>>

Available 2003-10-27

#### AHDS: Guides to Good Practice in the Creation and Use of Digital Resources

<<http://ahds.ac.uk/guides.htm>>

Available 2003-10-27

#### TASI

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

Available 2003-10-27

#### A Feasibility Study for the JISC Image Digitisation Initiative (JIDI)

<[http://heds.herts.ac.uk/resources/papers/jidi\\_fs.html](http://heds.herts.ac.uk/resources/papers/jidi_fs.html)>

#### Joint NPO and RLG Preservation Conference Guidelines for Digital Imaging 28 - 30 September 1998

<<http://www.rlg.org/preserv/joint/>>

Available 2003-10-27

#### Guides to Quality in Visual Resource Imaging:

<<http://www.rlg.ac.uk/visguides/>>

## 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:*

Joint NPO and RLG Preservation Conference Guidelines for Digital Imaging 28 - 30 September 1998

<<http://www.rlg.org/preserv/joint/>>

Available 2003-10-27

Preservation Management of Digital Materials

<<http://www.dpconline.org/graphics/handbook/>>

Available 2003-11-20

The Digitisation Process

<<http://www.ukoln.ac.uk/nof/support/help/papers/digitisation.htm>>

### 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 are outlined in 7.1.) 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

#### 5.1.1.1. 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.

For XHTML documents, the XML declaration **may** be omitted, but the encoding **must** be recorded within the value of the `http-equiv` attribute of a `meta` element.

For character encoding issues in the delivery of documents, see 7.1.1.1.

#### **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 2003-10-27

Extensible Markup Language (XML) 1.0

`<http://www.w3.org/TR/REC-xml/>`

Available 2003-10-27

XHTML 1.0 The Extensible HyperText Markup Language

`<http://www.w3.org/TR/xhtml1/>`

Available 2003-10-27

#### **Guidance:**

Jukka Korpela, A Tutorial on Character Code Issues

`<http://www.cs.tut.fi/~jkorpela/chars.html>`

Available 2003-10-27

#### 5.1.1.2. 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.

[Add reference to registries of appropriate schemas?]

In some instances, projects **may** choose to store text-based content using Adobe Portable Document Format (PDF). PDF is 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, as with any proprietary solution, there are dangers in its adoption and projects **should** be aware of the potential costs of this approach and **should** explore a migration strategy that will enable a future transition to open standards to be made. (See also section 7.1.1 for considerations regarding the accessibility of PDF documents).

#### ***Standards:***

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

Extensible Markup Language (XML) 1.0  
<<http://www.w3.org/TR/REC-xml/>>  
Available 2003-10-27

Text Encoding Initiative (TEI)  
<<http://www.tei-c.org/>>  
Available 2003-10-27

HTML 4.01 HyperText Markup Language  
<<http://www.w3.org/TR/html401/>>  
Available 2003-10-27

XHTML 1.0 The Extensible HyperText Markup Language  
<<http://www.w3.org/TR/xhtml1/>>  
Available 2003-10-27

#### ***Other references:***

Portable Document Format (PDF)  
<<http://www.adobe.com/products/acrobat/adobepdf.html>>  
Available 2003-10-27

#### ***Guidance:***

AHDS Guide to Good Practice: Creating and Documenting Electronic Texts  
<<http://ota.ahds.ac.uk/documents/creating/>>  
Available 2003-10-27

### 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.

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES  
WORKING DRAFT VERSION 0.06  
REVISED 23 DECEMBER 2003

### 5.1.2.1. *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. (Source: EMII DCF)

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. (Source: NOF-digi)

#### **Standards:**

Tagged Image File Format (TIFF)

<<http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf>>  
Available 2003-10-27

Joint Photographic Expert Group (JPEG)

<<http://www.w3.org/Graphics/JPEG/>>  
Available 2003-11-01

JPEG Still Picture Interchange File Format (SPIFF)  
<<http://www.jpeg.org/public/spiff.pdf>>  
Available 2003-11-15

**Guidance:**

TASI: Advice: Creating Digital Images  
<<http://www.tasi.ac.uk/advice/creating/creating.html>>  
Available 2003-11-15

*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. (N.B. Images resulting from the digitisation of physical line drawings should be managed as described in the previous section.)

**Standards:**

Portable Network Graphics (PNG)  
<<http://www.w3.org/TR/PNG>>  
Available 2003-10-27

5.1.2.2. *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 Flash format 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.

**Standards:**

Scalable Vector Graphics (SVG)  
<<http://www.w3.org/TR/SVG/>>  
Available 2003-10-27

**Other references:**

Macromedia Flash  
<<http://www.macromedia.com/>>  
Available 2003-10-27

### 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 AVI format, 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:**

Moving Pictures Experts Group (MPEG)  
<<http://www.cselt.it/mpeg/>>  
Available 2003-11-03

#### 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. 16-bit stereo sound at 44 KHz sample rate (CD standard) **should** be used for master copies.

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

## 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. (Source: Minerva GPH, DPC)

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. (Source: Minerva GPH, DPC)

Preservation Management of Digital Materials  
<<http://www.dpconline.org/graphics/handbook/>>  
Available 2003-11-20

TASI: Advice: Using CD-R and DVD-R for Digital Preservation  
<<http://www.tasi.ac.uk/advice/delivering/cdr-dvdr.html>>  
Available 2003-11-20

### 5.3. Migration 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. 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.

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. The capture of metadata is a critical part of a migration-based preservation strategy (see 6.2.3).

## 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

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES  
WORKING DRAFT VERSION 0.06  
REVISED 23 DECEMBER 2003

Getty Research Institute, Categories for the Description of Works of Art (CDWA)  
<[http://www.getty.edu/research/conducting\\_research/standards/cdwa/](http://www.getty.edu/research/conducting_research/standards/cdwa/)>  
Available 2003-11-15

International Standard for Archival Description (General) (ISAD(G)). Second Edition.  
<[http://www.ica.org/biblio/isad\\_g\\_2e.pdf](http://www.ica.org/biblio/isad_g_2e.pdf)>  
Available 2003-11-15

International Standard Archival Authority Record for Corporate Bodies, Persons and Families.  
<[http://www.ica.org/biblio/isaar\\_eng.pdf](http://www.ica.org/biblio/isaar_eng.pdf)>  
Available 2003-11-15

Encoded Archival Description (EAD)  
<<http://www.loc.gov/ead/>>  
Available 2003-11-15

Machine Readable Cataloguing (MARC): MARC 21  
<<http://www.loc.gov/marc/>>  
Available 2003-11-15

[should CIDOC CRM be here as well??]

**Guidance:**

Online Archive of California Best Practice Guidelines for Digital Objects (OAC BPG DO), Version 1.0  
<<http://www.oac.cdlib.org/oac-bpgdo/OAC-BPGDO-md1a.html>>  
Available 2003-11-15

### 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 2003-11-15

DC.Culture  
<<http://www.minervaeurope.org/DC.Culture.htm>>  
Available 2003-11-15

**Guidance:**

Using Dublin Core  
<<http://dublincore.org/documents/usageguide/>>  
Available 2003-11-15

## 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.

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

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 Data Dictionary for Technical Metadata for Digital Still Images

<<http://www.niso.org/pdfs/DataDict.pdf>>

Available 2003-11-15

RLG Working Group on Preservation Issues of Metadata

<<http://www.rlg.org/preserv/presmeta.html>>

Available 2003-11-15

### 6.2.3. Preservation Metadata

[in progress]

**Standards**

Reference Model for an Open Archival Information System (OAIS)

<<http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf>>

A Metadata Framework to Support the Preservation of Digital Objects

<[http://www.oclc.org/research/projects/pmwg/pm\\_framework.pdf](http://www.oclc.org/research/projects/pmwg/pm_framework.pdf)>

### 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 2003-11-17

IMS Content Packaging.

<<http://www.imsproject.org/content/packaging/>>

Available 2003-11-17

### 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 Research Support Libraries Programme (RSLP) Collection Description schema, the collection-level description schema define by Minerva, and the emerging Dublin Core Collection Description Application Profile.

#### **Standards:**

RSLP Collection Description

<<http://www.ukoln.ac.uk/metadata/rslp/>>

Available 2003-11-15

Minerva: Deliverable D3.1: Inventories, discovery of digitised content & multilingual issues: Feasibility survey of the common platform

<[http://www.minervaeurope.org/intranet/reports/D3\\_2.pdf](http://www.minervaeurope.org/intranet/reports/D3_2.pdf)>

Available 2003-11-15

Dublin Core Collection Description Application Profile

<<http://dublincore.org/groups/collections/>>

Available 2003-11-15

#### **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](http://www.minervaeurope.org/intranet/reports/D3_1.pdf)>

Available 2003-11-15

Collection Description Focus

<<http://www.ukoln.ac.uk/cd-focus/>>

Available 2003-11-15

## 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. 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 D3.1: Inventories, discovery of digitised content & multilingual issues: Feasibility survey of the common platform

<[http://www.minervaeurope.org/intranet/reports/D3\\_2.pdf](http://www.minervaeurope.org/intranet/reports/D3_2.pdf)>

Available 2003-11-15

## 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

##### 7.1.1.1. *Character Encoding*

The character encoding used in text-based documents **should** be transmitted in the HTTP header, and also recorded within documents as appropriate (see 5.1.1.1).

Note that some XML-based protocols may mandate the use of a specified character encoding, e.g. the OAI Protocol for Metadata Harvesting (see 8.1) 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 2003-10-27

##### 7.1.1.2. *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 (Source: NOF-digi).

In some cases, delivery in proprietary formats such as PDF, RTF or Microsoft Word **may** be appropriate as a supplementary format to XHTML/HTML, but projects **must** ensure that accessibility issues have been addressed (see 0) (Source: NOF-digi).

**Standards:**

HTML 4.01 HyperText Markup Language  
<<http://www.w3.org/TR/html401/>>  
Available 2003-10-27

XHTML 1.0 The Extensible HyperText Markup Language  
<<http://www.w3.org/TR/xhtml1/>>  
Available 2003-10-27

**Other references:**

Portable Document Format (PDF)  
<<http://www.adobe.com/products/acrobat/adobepdf.html>>  
Available 2003-10-27

## 7.1.2. Delivery of Still Images

### 7.1.2.1. *Photographic images*

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, with 120 pixels for the longest dimension as the preferred size. (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).

### 7.1.2.2. *Graphic non-vector images*

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

### 7.1.2.3. *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.

#### 7.1.3.1. *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. (Source: NOF-Digi, EMII-DCF)

Moving Pictures Experts Group (MPEG)  
<<http://www.cselt.it/mpeg/>>

#### 7.1.3.2. *Streaming*

Video for streaming **should** be delivered on the Web using Microsoft Advanced Streaming Format (ASF), Windows Media Video (WMV) or Apple Quicktime formats. (Source: NOF-Digi, EMII-DCF)

### 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.

#### 7.1.4.1. *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 (Source: EMII-DCF).

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

#### 7.1.4.2. *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.

### 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. (Source: NOF-digi, JISC IE)

Where appropriate, projects **may** wish to consider the use of 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/>>

Digital Object Identifier (DOI)  
<<http://www.doi.org/>>

## 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/>>

X3D  
<<http://www.web3d.org/x3d.html>>

QuickTime VR  
<<http://www.apple.com/quicktime/qtvr/>>

**Guidance:**

Archaeology Data Service VR Guide to Good Practice  
<<http://ads.ahds.ac.uk/project/goodguides/g2gp.html>>

## 7.3. Geographic Information Systems

[Need more about use of GIS]

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES  
WORKING DRAFT VERSION 0.06  
REVISED 23 DECEMBER 2003

Projects seeking to employ Geographic Information Systems (GIS) **must** obtain appropriate permissions for use of any map data from third parties, ensuring that licences extend to delivering services to the public over the Internet.

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 an appropriate standard co-ordinate reference system when recording spatial data.

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

**Standards:**

OpenGIS Consortium  
<<http://www.opengis.org/>>

**Guidance:**

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

## 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. If other protocols are used (e.g. Z39.50) gateways **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  
<<http://www.w3.org/Protocols/HTTP/>>

### 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 and hardware devices (e.g. Personal Digital Assistants (PDAs) as well as PCs). 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. Projects **must** achieve WAI level A conformance; projects **should** aim to achieve WAI level AA conformance.

**Standards:**

Cascading Style Sheets (CSS), Level 2  
<<http://www.w3.org/TR/REC-CSS2/>>  
Available 2003-10-27

Web Content Accessibility Guidelines (WCAG) 1.0  
<<http://www.w3.org/TR/WCAG10/>>  
Available 2003-10-27

**Guidance:**

Web Accessibility Initiative (WAI)  
<<http://www.w3.org/WAI/>>  
Available 2003-10-27

RNIB: Accessible Web Design  
<<http://www.rnib.org.uk/digital/hints.htm>>  
Available 2003-11-15

Watchfire Bobby Online Service  
<<http://bobby.watchfire.com/>>  
Available 2003-11-15

Jakob Nielsen, useit.com  
<<http://www.useit.com/>>  
Available 2003-11-15

## 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:**

ISO 17799  
(which site?!?)

Secure Sockets Layer (SSL) 3.0  
<<http://wp.netscape.com/eng/ssl3/>>

**Guidance:**

Introduction to Secure Socket Layer (SSL)  
<<http://developer.netscape.com/docs/manuals/security/sslin/index.htm>>

### 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.

Museums providing Web sites **should** consider registering a “.museum” top-level domain name in order to improve disclosure of their services and to indicate that their sites are associated with genuine museums.

**Guidance:**

Domain Names System (DNS) Resources Directory  
<<http://www.dns.net/dnsrd/>>

Dot Museum  
<<http://about.museum/>>

### 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  
<<http://www.w3.org/Protocols/HTTP/>>

#### 7.4.5. Performance Indicators

Performance indicators can be used to provide objective measures of the usage of a Web service and provide some indication of the impact of the digitisation project. The most popular performance indicator makes use of Web server log files. Analysis of server log files can provide valuable information on the growth of a service and usage patterns, although reports need to be interpreted carefully.

Projects **must** maintain statistics about the usage of Web sites and **should** use them appropriately to analyse the usage of the digitised resources.

Further guidance in this area will be developed.

## 8. Disclosure 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.

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 DC.Culture schema (See 6.2.1).

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 2003-11-15

DC.Culture

<<http://www.minervaeurope.org/DC.Culture.htm>>

Available 2003-11-15

IEEE Learning Object Metadata

<<http://ltsc.ieee.org/wg12/>>

Available 2003-11-15

### 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 2003-11-17

**Guidance:**

OAI FAQs

<<http://www.ukoln.ac.uk/distributed-systems/dner/arch/faq/oai/>>

## 8.2. Distributed Searching

Projects **may** need to display awareness of Z39.50, a network protocol that allows searching of (usually remote) heterogeneous databases and retrieval of data, via one user interface. 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.

Projects **may** also 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.

**Standards:**

Z39.50 Maintenance Agency

<<http://www.loc.gov/z3950/agency/>>

Available 2003-11-17

Bath Profile

<<http://www.nlc-bnc.ca/bath/tp-bath2-e.htm>>

Available 2003-11-17

SRW: Search/Retrieve Web Service

<<http://lcweb.loc.gov/z3950/agency/zing/srw/>>

Available 2003-11-17

**Guidance:**

Z39.50 for All

<<http://www.ariadne.ac.uk/issue21/z3950/>>

Available 2003-11-17

## 8.3. Alerting

Projects **may** need to demonstrate awareness of the RDF (or Rich) Site Summary (RSS) family of specifications. 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.

***Standards:***

RDF Site Summary (RSS) 1.0

<<http://purl.org/rss/1.0/spec>>

Available 2003-11-17

RSS 2.0

<<http://blogs.law.harvard.edu/tech/rss>>

Available 2003-11-17

***Guidance:***

Syndicated content: it's more than just some file formats

<<http://www.ariadne.ac.uk/issue35/miller/>>

Available 2003-11-17

## 9. Re-use and re-purposing

Users will want to repackage and re-purpose material that has been developed by digitisation projects. 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 (See 8).

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 2003-11-17

IMS Global Learning Consortium, Inc.

`<http://www.imsproject.org/>`

Available 2003-11-17

IMS Content Packaging.

`<http://www.imsproject.org/content/packaging/>`

Available 2003-11-17

## 10. IPR and Copyright

[This section to be revised in light of EMII-DCF IPR material. Also, include Creative Commons]

### 10.1. Establishing 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.

Projects **should** record the existence of rights held in and permissions granted for all resources.

[Reference Creative Commons in 11.1 or 11.2??]

[Need more generic references here]

Copyright and the Networked Environment – Issue Paper from the Networked Services Policy Taskgroup

<<http://www.earl.org.uk/policy/issuepapers/copyright.html>>



Creating Digital Resources for the Visual Arts: Standards and Good Practice  
<[http://vads.ahds.ac.uk/guides/creating\\_guide/contents.html](http://vads.ahds.ac.uk/guides/creating_guide/contents.html)>

JISC Management Briefing Paper on Copyright  
<[http://www.jisc.ac.uk/pub98/sm05\\_copyright.html](http://www.jisc.ac.uk/pub98/sm05_copyright.html)>

UK Intellectual Property  
<<http://www.intellectual-property.gov.uk>>

World Intellectual Property Organization  
<<http://www.wipo.org/>>

## 10.2. Safeguarding Copyright

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. 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

### 10.2.2. 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/>>