# Technical Guidelines for Digital Cultural Content Creation Programmes

*Working Draft Version 0.01: Revised 15 September 2003*

TECHNICAL  GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

1

# Acknowledgements

This document is based on the *NOF-digitise Technical Standards and Guidelines* (Version 5, February 2003), that was 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.

NOF-digitise Technical Standards and Guidelines
<http://www.peoplesnetwork.gov.uk/content/technical.asp>

This current document draws on the structure and a large part of the content of that document, and incorporates various updates and revisions that reflect technological (and other) developments that have occurred since the launch of the NOF-digitise programme in 1999. It also seeks to focus on the provision of guidelines that will be applicable across a range of similar contexts.

It also draws on a number of other sources:

NOF-digi Technical Advisory Service FAQ
<http://www.ukoln.ac.uk/nof/support/help/faqs/>

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>

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

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

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>

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

The initial NOF-digitise standards and guidelines drew on *Scientific, Industrial and Cultural Heritage: a shared approach*, a research framework for digital libraries, museums and archives written by Lorcan Dempsey for the European Commission's Information Society Directorate General.

Scientific, Industrial and Cultural Heritage: a shared approach
<http://www.ariadne.ac.uk/issue22/dempsey/>

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

2

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

3

# Table of Contents

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

4

TECHNICAL  GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

5

TECHNICAL  GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

6

# 1. Introduction

Throughout Europe, international, national, regional and local initiatives are investing significant public and private sector funding in making a range of cultural heritage resources available through digital channels. The motivations and drivers for these initiatives may vary widely: they may encompass different classes 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 resultant content 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'.

## 1.1. The Role of Technical Standards

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. This document seeks to provide some guidelines for the use of standards - primarily technical standards - by projects operating within programmes for the creation of digital cultural content.

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. 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 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.

Important areas for consideration include:

- **Interoperability**. It is important that content can be shared seamlessly between projects and users, and between projects – both within and beyond the scope of a single content creation programme. 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 widely useful resource it may be necessary to consider support for multiple language communities and accessibility to citizens with 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

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

7

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:

- **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, and for content that fails to reach users or markets.

- **Creators, authors**. Their legacy to the future may be lost.

It is also recognised that a programme may cover projects with diverse requirements and with specific questions not fully addressed by programme-level guidelines. Further, even within the lifetime of a programme, the technological environment changes and standards evolve. It is common for programmes to 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 Life Cycle Approach

A 'life cycle' approach has been used in the discussion of standards in this document. Decisions taken at any stage in a resource's life may have ramifications for cost-effective use at other stages.

Also, it is emphasised that a digital resource is created not in isolation but as part of a digital collection, and the life cycle of the resource should be considered within the context of that collection and the development of the collection. Indeed, collections themselves are no longer considered as self-contained entities; rather they are seen as components around which many different types of digital services might be constructed.

For this reason the development of a collection is highlighted as a distinct 'stage' in the resource life cycle:

- **Collection Development**. The digital resource will typically be created as part of a collection of digital resources. A collection requires management; it should also be capable of development.

- **Creation**. The actual creation of an individual digital resource.

- **Management**. The digital resource must be managed in some way if it is to be accessible and meaningful through time.

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

8

- **Access**. Materials need to be made available on the network in accessible, usable, secure, and responsible ways.

- **Reuse and repackaging**. A digital resource may be used in more than one way. For example, a digitised local history photograph, created initially as part of an online exhibition, may be repackaged as part of a learning resource.

## 1.3. 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 the NOF-digitise standards on which this document is based 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.

This current document does not include this distinction between requirements and guidance on the grounds that distinction 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 **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.  The NOF-digitise standards document describes its use of these terms as follows:

- **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 should 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

TECHNICAL  GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

9

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/>`

## 1.4.  Other Standards

The guidelines provided by this document are intended to be generally applicable. Each digitisation 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).

Programmes **should** maintain awareness of all ongoing standards developments relevant to their operating context.

TECHNICAL  GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

10

# 2. Collection Development

## 2.1.  Collection Development Policy

A digital collection is not simply a random assemblage of digital objects. A collection is a managed resource that is the result of a process of selection and organisation. Projects **should** be able to identify audiences for their collection and to understand the requirements of those audiences, taking into account the wider context of the availability of significant related resources. All projects **should** establish a collection development policy before beginning to digitise objects.

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

## 2.2.  Collection-Level Description

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** provide collection-level descriptions using an appropriate metadata schema.  Projects **should** display awareness of the Research Support Libraries Programme (RSLP) Collection Description schema and the emerging Dublin Core Collection Description Application Profile.

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

DC Collection Description Application Profile
<http://dublincore.org/groups/collections/>

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

## 2.3.  Collection Sustainability

The material digitised by each project **should** be developed as a digital collection that will continue to grow and be added to beyond the lifetime of the project.

Where a collection is created with specific funding, a project **should** have a plan for ensuring the continued accessibility, usability and development of the collection beyond the period of that initial funding.

TECHNICAL  GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

11

# 3. Creation

This section covers the actual creation of digital resources by projects. Technical standards decisions made at this stage will have a fundamental impact on the manageability, accessibility and viability of the resources created.

The following section on file format requirements for creation and storage of resources (3.1) should be read in conjunction with the recommendations for file formats for delivery of resources (5.1).

## 3.1.  File Formats

Open standard formats **should** be used when creating digital records in order to maximise access to resources. (Note that file formats for the delivery of digital records to users are outlined in 5.1.) 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 standards **may** be acceptable. However, where proprietary standards 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 standards within their proposals for funding, paying particular attention to issues of accessibility.

The Diffuse project provides a single entry point to up-to-date reference and guidance information on available and emerging standards and specifications that facilitate the electronic exchange of information.

Diffuse
`<http://www.diffuse.org>`

### 3.1.1.  Text Formats

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

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 **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). All such documents **should** be validated against a published DTD or XML Schema.

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**

TECHNICAL  GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

12

be omitted, but the encoding **must** appear within the value of the `http-equiv` attribute of a `meta` element.

Projects **may** choose to store such content either in plain files or within a database of some kind.

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** explore a migration strategy that will enable a future transition to open standards to be made. They **must** also ensure that accessibility issues have been addressed – Jakob Nielsen's Web site is a useful source of information about these issues.

AHDS Guide to Good Practice: Creating and Documenting Electronic Texts
`<http://ota.ahds.ac.uk/documents/creating/>`

Text Encoding Initiative (TEI)
`<http://www.tei-c.org/>`

W3C HTML
`<http://www.w3.org/MarkUp/>`

The Unicode Standard - A Technical Introduction
`<http://www.unicode.org/unicode/standard/principles.html>`

PDF
`<http://www.adobe.com/products/ acrobat/adobepdf.html>`

useit.com: Jakob Nielsen's Web site
`<http://www.useit.com/>`

## 3.1.2. Image, Video and Audio formats

Images, video, sound and music should initially be created and stored using appropriate open or proprietary standard formats as outlined below.

- Images **must** be created using one of the following formats – TIFF, PNG, GIF or JPEG/SPIFF. In general photographic images **should** be created as TIFF images. However, in cases, for example when using cheaper digital cameras, it may be appropriate to use JPEG/SPIFF as an alternative. 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. Line drawings, such as certain types of computer-generated imagery, **should** normally be created as PNG or GIF images.

- Vector graphics **should** be created and stored using a rich open format such as the SVG (Scalable Vector Graphics) format. 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.

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

13

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

- Sound **must** be created and stored using MP3, Microsoft WAV or WMA, RealAudio, or Sun AU formats.

SVG
<http://www.w3.org/TR/SVG/>

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

## 3.1.3. 3D Content

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, 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 **should** test them using typical modem connections and home, school, or library computer systems.

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.

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

Web3D Consortium
<http://www.web3d.org/>

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

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

TECHNICAL  GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

14

## 3.2. Geographic Information Systems

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.

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

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

## 3.3. Data Capture

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 Feasibility Study for the JISC Image Digitisation Initiative (JIDI)
`<http://heds.herts.ac.uk/resources/papers/jidi_fs.html>`

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

AHDS: Guides to Good Practice in the Creation and Use of Digital Resources
`<http://ahds.ac.uk/guides.htm>`

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

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

15

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.

Joint NPO and RLG Preservation Conference Guidelines for Digital Imaging 28 - 30 September 1998
`<http://www.rlg.org/preserv/joint/>`

Guides to Quality in Visual Resource Imaging:
`<http://www.rlg.ac.uk/visguides/>`

## 3.4. Metadata

Metadata is structured data about data. Metadata is created to support a range of operations and the term typically refers to any data used to aid the identification, description, location or use of resources. Metadata is required through the whole life cycle of a digital resource and is consequently referred to throughout this document.

Projects **may** need to record metadata about both the digital objects created during the digitisation process and (if appropriate) about physical objects from which those digital objects were derived.

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.

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`

### 3.4.1. Descriptive Metadata

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

The Dublin Core Metadata Element Set (DCMES) is an example of a very simple descriptive metadata schema. 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.

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

Projects **must** capture and store sufficient descriptive metadata to be able to generate a simple/unqualified Dublin Core metadata record for each item.

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

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

16

Dublin Core Metadata Element Set
http://dublincore.org/documents/dces/

Using Dublin Core
http://dublincore.org/documents/usageguide/

### 3.4.2. 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;

- Rights management metadata, describing copyright, use restrictions and license agreements that constrain the use of the 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.

In many cases, this 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. In some cases 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 (July 2000).

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.

NISO Data Dictionary for Technical Metadata for Digital Still Images
<http://www.niso.org/DataDict.html>

RLG Working Group on Preservation Issues of Metadata
<http://www.rlg.org/preserv/presmeta.html>

### 3.4.3. 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.

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

17

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.

Projects **should** show understanding of the requirements for structural metadata for digital objects, and of the role of METS in "wrapping" metadata and digital objects.

Metadata Encoding and Transmission Standard (METS)
`<http://www.loc.gov/standards/mets/>`

## 3.5.  Preservation

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.  More guidance on this topic can be found in Guidelines for digital imaging conference papers.

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 this – in many cases – can only be done 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 accessing via the Web) can be derived.  This will help with the periodic migration of data and with the development of new products and resources.

Guidelines for digital imaging: papers
`<http://www.rlg.org/preserv/joint/>`

Preservation Management of Digital Materials
`<http://www.jisc.ac.uk/dner/preservation/workbook/>`

TECHNICAL  GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

18

# 4. Management

In the same way that physical collections are managed, digital collections also need to be managed.  Issues to be addressed include preservation of the data, copyright issues, security and e-commerce.  Again, standards decisions made at this stage will have long term implications for the accessibility and usability of the data.

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

Projects **should** make use of the Uniform Resource Identifier (URI) for this purpose.

Projects **may** wish to consider assigning Digital Object Identifiers to resources where appropriate.

Uniform Resource Identifiers
`<http://www.w3.org/Addressing/>`

Digital Object Identifier
`<http://www.doi.org/>`

## 4.2.  Intellectual Property Rights

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

TECHNICAL  GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

19

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.

[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>

JISC Management Briefing Paper on Copyright
<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/>

## 4.3.  Performance Indicators

Performance indicators can be used to provide objective measures of the usage of a Web service.  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.

Web projects **should** maintain server log files and **should** use them appropriately to analyse the usage of the digitised resources.

## 4.4.  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.

TECHNICAL  GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

20

Projects **should** be managed in accordance with 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.

ISO 17799
(which site?!!)

Introduction to SSL
<http://developer.netscape.com/docs/manuals/
security/sslin/index.htm>

## 4.5. E-Commerce

[Is this section required?]

## 4.6. Preservation

Once a digital surrogate has been created, its own preservation then becomes an issue. There is (as yet) no standardised way of ensuring the long-term preservation of any digital information, but all projects **should** be aware of, and demonstrate that they are aware of the need for preserving their digital resources. Decisions made at the creation stage will have effects further down a digital object's life cycle.

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. That said, 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.

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

## 4.7. Rights Metadata

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.

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

21

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.

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

22

# 5. Access

It is expected that end-user access will be through the use of Internet Protocols.

## 5.1. Access to Resources

Projects **must** be accessible using a Web browser. This will normally be achieved using HTML or XHTML and the HTTP 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.

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

Web services **must** be accessible to a wide range of browsers and hardware devices (e.g. Personal Digital Assistants (PDAs) as well as PCs). Web services **must** be usable by browsers that support W3C recommendations such as HTML, 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.

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 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 Recommendations and so ensure a high degree of accessibility for people with disabilities. Projects **should** achieve WAI level A (or level 1) accessibility compliance.

The following recommendations on file formats should be read in conjunction with the requirements for file formats for storage of resources (3.1).

### 5.1.1. Text formats

Text-based content **must** be delivered as XHTML 1.0 or HTML 4 in most cases, though the use of XML with other DTDs **may** sometimes be appropriate.

The character encoding used in text-based documents **should** be declared in the HTTP header, and also within documents as appropriate (see 3.1.1).

In some cases, delivery in proprietary formats such as PDF, RTF or Microsoft Word **may** be appropriate. In these cases, projects **must** justify their requirement for use of proprietary standards within their business plan and **must** explore a migration strategy that will enable a transition to open standards to be made in the future.

### 5.1.2. Image, video and audio formats

Images, video and sound **must** be delivered using appropriate open standard formats or the proprietary formats as outlined below.

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

23

- Images **must** be provided on the Web as GIF (for line-drawings) or JPEG/SPIFF (for photographs) formats. PNG **may** be an alternative format to GIF.

- Video **must** be delivered using MPEG or Microsoft AVI, ASF, WMF or Quicktime formats.

- Sound **must** be delivered using proprietary MP3, RealAudio, Microsoft WAV, WMA or Sun AU formats.

Hypertext Transfer Protocol, HTTP/1.1
`<http://www.w3.org/Protocols/HTTP/>`

W3C WAI recommendations
`<http://www.w3.org/WAI/>`

RNIB: Accessible Web Design
`<http://www.rnib.org.uk/digital/hints.htm>`

Bobby Accessibility Approval Service
`<http://www.cast.org/bobby/>`

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

## 5.2. 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.

DNS Resources Directory
`<http://www.dns.net/dnsrd/>`

## 5.3. Metadata and Resource Discovery

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 **should** be capable of delivering:

- a collection-level metadata record describing the collection as a whole;

- item-level metadata records describing individual digital resources within the collection.

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

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

24

In order to facilitate potential exchange and interoperability between services, projects **must** be able to provide item level descriptions in the form of simple, unqualified Dublin Core metadata.  (See 3.4.1)

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

## 5.3.1.  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 rights statement in their metadata to ensure they retain ownership rights in their metadata.

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

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

## 5.3.2.  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 **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.

Z39.50 Maintenance Agency
`<http://www.loc.gov/z3950/agency/>`

Z39.50 for All
`<http://www.ariadne.ac.uk/issue21/z3950/>`

SRW: Search/Retrieve Web Service
`<http://lcweb.loc.gov/z3950/agency/zing/srw/>`

## 5.3.3.  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.

TECHNICAL  GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

25

RDF Site Summary (RSS) 1.0
<http://purl.org/rss/1.0/spec>

RSS 2.0
<http://blogs.law.harvard.edu/tech/rss>

Syndicated content: it's more than just some file formats
<http://www.ariadne.ac.uk/issue35/miller/>

## 5.4. 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.

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

## 5.5. 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.

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

26

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

Hypertext Transfer Protocol, HTTP/1.1
`<http://www.w3.org/Protocols/HTTP/>`

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

27

# 6. Re-use

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.

## 6.1. 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 **should** demonstrate awareness of the IEEE Learning Object Metadata standard. Project **should** track the work of the IMS consortium in developing specifications to support interoperability amongst learning technology systems.

IEEE Learning Object Metadata
`<http://ltsc.ieee.org/wg12/>`

IMS: Global Learning Consortium, Inc.
`<http://www.imsproject.org/>`

TECHNICAL GUIDELINES FOR DIGITAL CULTURAL CONTENT CREATION PROGRAMMES
WORKING DRAFT VERSION 0.01
REVISED 24 SEPTEMBER 2003

28