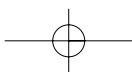
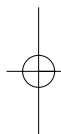
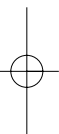


Technical Guidelines for Digital Cultural Content Creation Programmes



Technical Guidelines for Digital Cultural Content Creation Programmes

Version 2.0: September 2008

General co-ordination

Rossella Caffo (MinervaEC Project Manager)

Antonella Fresa (MinervaEC Technical Coordinator)

Pier Giacomo Sola (MinervaEC Organisation Manager)

Editorial committee

Version 2.0 was edited by Kate Fernie (Consultant, UK) with the assistance of Giuliana De Francesco (Ministero per i Beni e le Attività Culturali, IT) and David Dawson (Wiltshire Heritage Museum, UK) for the MINERVA-eC project.

Version 1.0 of this document was edited by Pete Johnston (UKOLN).

Web version

`<http://www.minervaeurope.org/publications/technicalguidelines.htm>`

Design

due_pavese

Geo Graphic sdf

Contributors

The contributors to Version 2.0 include: Brian Kelly, UKOLN (UK); Nick Poole, Collections Trust (UK); Tom Morgan, Museum Copyright Group and National Portrait Gallery (UK); Maria Teresa Natale, MinervaEC project (IT); Susan Hazan, Israel Museum (IS); Mary Rowlatt, consultant (UK); Rob Davies, MDR Partners / EuropeanaLocal project (UK); Stefanos Kollias, National Technical University of Athens (GR); Wendy Sudbury (UK); Andy Powell, Eduserve (UK); Karla Youngs, TASI (UK); Paolo Auer, Ente per le nuove tecnologie, l'energia e l'ambiente ENEA(IT); Stefano Casati, Istituto e museo di storia della scienza(IT); Andrea D'Andrea, Università di Napoli L'Orientale (IT); Pierluigi Feliciati, Università d Macerata (IT); Francesca Klein, Archivio di Stato di Firenze (IT); Franco Lotti, Consiglio nazionale delle ricerche, Istituto di fisica applicata "Nello Carrara" (IT); Cristina Magliano, Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (IT); Gianna Megli, Biblioteca Nazionale Centrale di Firenze (IT); Anna Maria Tammaro, Università di Parma (IT); Milena Dobрева, University of Strathclyde (UK).

The contributors to Version 1.0 of this document included: Eelco Bruinsma, Consultant, NL; Rob Davies, MDR Partners / PULMAN Project, UK; David Dawson, MLA, UK; Bert Degenhard Drenth, Adlib Information Systems / EMII-DCF Project, NL; Giuliana De Francesco, Ministero per i beni e le attività culturali, IT; Muriel Foulonneau, Relais Culture Europe / EMII-DCF Project, FR; Gordon McKenna, mda / EMII-DCF project, UK; Paul Miller, UKOLN, UK; Maureen Potter, ERPANET Project, NL; Jos Taekema, Digital Erfgoed Nederland, NL and Chris Turner, MLA, UK.

This document is based primarily on four sources.

- the *NOF-digitise Technical Standards and Guidelines* (Version 5, February 2003), that were developed on behalf of the UK New Opportunities Fund (NOF), by UKOLN, University of Bath, in association with Resource: The Council for Museums, Archives & Libraries (now known as MLA). <http://www.mla.gov.uk/resources/assets//T/technicalstandardsv1_pdf_79b4.pdf>.
- additional information provided to NOF-digitise projects in support of the Standards and Guidelines by the NOF-digitise Technical Advisory Service, operated for NOF by UKOLN and the Arts and Humanities Data Service (AHDS), in the form of the programme manual, briefing papers and FAQs. <<http://www.ukoln.ac.uk/nof/support/manual/>> and <<http://www.ukoln.ac.uk/nof/support/help/faqs/>>.
- the *Framework Report* (September 2003), published by the European Museums' Information Institute – Distributed Content Framework (EMII-DCF) project, particularly the Data Capture Model in Chapter 16. <<http://www.emii.org/dcf-frame.pdf>>.
- the *Good Practice Handbook* (Version 1.2, November 2003), developed by the Minerva project (Working Group 6). <http://www.minervaeurope.org/structure/workinggroups/goodpract/document/bestpracticehandbookv1_2.pdf>.
- DCC Diffuse standards registry, <<http://www.dcc.ac.uk/diffuse/>>.

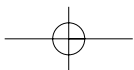
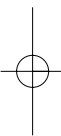
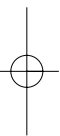
It also draws on a number of other sources:

- The Institute of Museum and Library Services' Framework of Guidance for Building Good Digital Collections <<http://www.niso.org/publications/rp/framework3.pdf>>
- 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 Digitisation Projects <http://www.pch.gc.ca/progs/pcce-ccop/pubs/ccop-pcceguide_e.pdf>
- Working with the Distributed National Electronic Resource (DNER): Standards and Guidelines to Build a National Resource <http://www.jisc.ac.uk/index.cfm?name=projman_standards>
- JISC Information Environment Architecture Standards Framework <<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/standards/>>
- The Public Libraries Managing Advanced Networks (PULMAN) Guidelines <<http://www.pulmanweb.org/DGMs/DGMs.htm>>

table of contents

Acknowledgements	5
1. Introduction	9
1.1. The Purpose of this Document	10
1.2. The Role of Technical Standards	11
1.3. The Benefits of Deploying Open Standards	12
1.4. The Life Cycle Approach	14
1.5. Requirement Levels	14
1.6. Summary	15
2. Projects and Planning	17
2.1. Introduction	17
2.2. Project Phases	18
2.3. Planning	20
2.4. People and Roles	21
2.6. Managing Risks	22
3. Preparing for the digitisation process	23
3.1. Selecting Materials for Digitisation	24
3.2. Preparing Original Materials	25
3.3. Staff Training	26
3.4. In-house Digitisation or Out-sourcing?	26
3.5. Hardware and Software	27
3.6. The Digitisation Process	29
4. Storage and Management of the Digital Master Material	31
4.1. File Formats	32
4.1.1. Text Capture and Storage	32
4.1.2. Still Image Capture and Storage	35
4.1.3. Video Capture and Storage	38
4.1.4. Audio Capture and Storage	38
4.1.5. Multimedia	39
4.1.6. GIS	40
4.1.7. 3D and Virtual Reality	40
4.2. Media Choices	44
4.3. Preservation Strategies	45
5. Metadata, standards and resource discovery	47
5.1. Metadata Standards	47
5.2.1. Descriptive Metadata	49
5.2.2. Administrative Metadata	50
5.2.3. Preservation Metadata	51
5.2.4. Structural Metadata	52
5.2.5. Collection-Level Description	53

5.3.	Terms and Conditions of Use	54
5.4.	Terminology Standards	54
5.4.	Resource Discovery	55
5.5.	Metadata, Ontologies and the Semantic Web	57
5.6.	OAI-PMH and Metadata Harvesting	59
5.7.	Distributed Searching	60
5.8.	Syndication / Alerting	61
6.	Publishing on the Web	63
6.2.	Browsers and Protocols	64
6.3.	Accessibility	64
6.4.	Security	65
6.5.	Authenticity	66
6.6.	User Authentication	66
6.7.	Search Engine Optimisation	67
6.8.	Current and Future Trends: Web 2.0-3.0	68
6.8.1	Technologies	69
6.8.2.	User perspectives	70
6.8.3	Factors to Take into Consideration	70
7.	Delivery formats	73
7.1.	Identification	73
7.2.	Delivery of Text	74
7.3.	Delivery of Still Images	76
7.4.	Delivery of Video	76
7.5.	Delivery of Audio	77
7.6.	Delivery of Virtual Reality	78
7.7.	Delivering Geographic Information	79
8.	Reuse and Re-purposing	81
8.1.	Learning Resource Creation and Re-use	81
9.	Intellectual Property Rights, Copyright, Licencing and Sustainability	83
9.1.	Identifying, Recording and Managing Intellectual Property Rights	84
9.2.	Safeguarding Intellectual Property Rights	85
9.2.1.	Creative Commons	85
9.2.2.	Planning for Sustainability	86
9.2.3.	Watermarking and Fingerprinting	86
	Appendix 1. About This Document	87
	Appendix 2. Glossary	89
	Appendix 3. Business models: Web 2.0-3.0	91



1. Introduction

These Technical Guidelines have been prepared in the context of a series of European and national initiatives in recent years to make digital content in Europe more accessible, usable and exploitable. The broader context for these Guidelines is the European Commission's i2010 Digital Libraries initiative, which has as one of its objectives the creation of a European Digital Library aiming to make millions of digital objects easily accessible to all European citizens. At national level, modernisation in public service delivery and the economy are major drivers for investment in digitisation and ICT.

The growth and development of new technologies is setting an exciting agenda for innovation and cultural institutions are increasingly aware of the potential. Growing quantities of digital content and information are becoming available in increasingly sophisticated forms. The emergence of web-based communities and services (such as social-networking sites), mobile and Wi-Fi technology offer new opportunities for citizens to create and share personal media and for cultural institutions to interact with their audiences. Both individuals and organisations in today's society are confronted with growing quantities of content and increasing demands for knowledge and skills. This requires that multilingual content and information be made more accessible and usable over time by humans and machines alike. The success of the European Digital Library initiative depends in part on the ability to unlock its users' abilities to access, manipulate and use cultural heritage resources. Across Europe, international, national, regional and local initiatives are investing significant public and private sector funds in digitisation. The motivations and drivers for these initiatives vary widely but typically all funding programmes aim to maximise their impact by requiring that the digital content produced is as widely useful, portable and durable as possible – in other words resources and content should be 'interoperable'.

Ensuring 'interoperability' involves:-

- consistency of approach to the creation, management and delivery of digital resources through the effective use of standards, the rules and good practice guidelines;

1. Introduction

- making content available to a range of services through the use of internet protocols and APIs (Application Programming Interfaces).

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

1.1 The Purpose of this Document

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 we recognise that different programmes will take different approaches to conformance with guidelines. This document seeks to identify areas where there is already a commonality of approach and to provide a core around which context-specific requirements might be built. The scope and emphasis is similar to that of the EMII-DCF *Data Capture Model* and several of the recommendations in this document are based directly on those presented in that model. Usage of these guidelines cannot guarantee 'interoperability': the precise requirements for usefulness, portability and durability of digital resources vary from programme to programme and the form in which standards are deployed by individual projects will reflect those requirements. While the guidelines provided by this document are intended to be generally applicable, each programme operates in a context and funded 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 often fall within the scope of standards mandated by national governments.

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 will be important to provide an advisory service for funded projects to offer guidance on the interpretation and implementation of standards and guidelines, and to update recommendations to reflect significant developments.

Those developing interoperable services should provide guidance and support for projects that are making their content available, recognising that standards can be implemented in a number of different ways.

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".

Appropriate use of standards in digitisation can deliver the consistency that makes interoperability possible. It means that a service operating across resources from multiple providers only needs to handle a limited number of clearly specified formats, interfaces and protocols. An ever-increasing number of different formats and protocols would make such a development complex, costly and at best unreliable, if not impossible. The process of developing standards also means that they capture good practice based on past experience and bring rigour to current practice.

Standards are often defined as either:

- *de jure* – formally recognized by a body responsible for setting and disseminating standards, usually developed through the collaboration of a number of interested parties. Examples are standards such as the TCP/IP set of protocols, maintained by the Internet Engineering Task Force (IETF) or the Adobe Portable Document Format 1.7 (PDF) which is now maintained by the ISO standards body.
- *de facto* – not formally recognized by a standards body but 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 AutoDesk's DWG format for CAD files or Windows.

The "openness" of a standard is a further consideration. This can refer to a number of characteristics of a standard. The EMII-DCF *Framework Report* (<<http://www.emii.org/dcf.htm>>) 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, for example through licensing); and

1. Introduction

- on-going support (driven by requirements of the user not the interests of the standard provider).

With an open standard the specifications of the formats, interfaces and protocols used by resource providers are openly available and multiple developers can develop similar tools and services. This means that dependency on a single tool or platform can be avoided.

The formal processes associated with the development of *de jure* standards are generally regarded as ensuring their "openness".

These guidelines give preference to open standards, but in some cases industry or *de facto* standards are also considered.

1.3 The Benefits of Deploying Open Standards

Important areas for consideration include:

- **Interoperability.** Content can be accessed seamlessly by users, across projects, across services 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.** Materials are as accessible as possible and are made publicly available using open standards and non-proprietary formats. Consideration is given to support for multiple language communities and ensure accessibility for citizens with a range of disabilities.
- **Preservation.** The long-term future of materials is secured, so that the benefit of the investment is maximized 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 and result in the waste of resources by:

- **Users - the citizen, the learner, the child.** They will waste time and effort as they cannot readily find or use the most appropriate resources for their needs, because they are not adequately described, or delivered in a particular way, or require specialist tools to exploit, or were not captured in a usable form.

1. Introduction

- **Information providers and managers.** Their investment may be redundant and wasted as their resources fail to be sufficiently used, their products reaching only part of the relevant audience, as they invested in non-standard or outmoded practices.
- **Service providers.** They will spend additional time and effort in implementing interoperable services, to link content together from different content providers where each project has implemented standards in different ways.
- **Funding agencies.** They have to pay for redundant, fragmented effort, for the unnecessary repetition of learning processes, for projects that operate less efficiently than they should and deploy techniques that are less than optimal, for content that fails to meet user needs or does not meet market requirements.
- **Creators, authors.** Their legacy to the future may be lost.

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

- **Complexity.** The process for developing standards may result in over-complex standards being developed, which may be difficult to exploit.
- **Failure to gain marketplace acceptance.** Standards may fail to gain acceptance in the market place, with a failure to develop significant numbers of tools which exploit the standard.
- **Costs.** The costs of deploying open standards may be too expensive for their intended use.
- **Lack of user interest.** Users may fail to make use of services which are based on open standards, preferring to continue to use services based on proprietary solutions.
- **Enhancement to proprietary solutions.** The owners of proprietary formats may respond to the challenges posed by open standards by making their formats more open, reducing the costs associated with their use of enhancing the functionality of their formats.
- A paper on "*A Contextual Framework For Standards*" expands on the challenges in selecting open standards and describes a framework to assist in the selection of appropriate open standards. See <http://www.ukoln.ac.uk/web-focus/papers/e-gov-workshop-2006/>.

In the light of these issues the approach which needs to be taken by cultural heritage organisations can be summarised as "institutions **should**

1. Introduction

seek to make use of open standards, **should** make their content available to interoperable services, and **must** have policies and procedures in place which documents their approaches and decisions related to their use of open or closed standards”.

1.4 The Life Cycle Approach

This structure of this document reflects a ‘life cycle’ approach to the digitisation process and (with some modifications) parallels the structure of the *Good Practice Handbook* developed by 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.

- Project Planning
- Preparing for the Digitisation Process
- Storage and Management of the Digital Master Material
- Metadata, standards and resource discovery
- Publishing on the Web
- Delivery formats
- Re-use and re-purposing
- Intellectual Property and Copyright

1.5 Requirement Levels

The approaches taken to conformance to standards and guidelines vary between programmes, ranging from encouraging the adoption of good practice to mandating conformance to standards as a condition of a grant award. Typically there are 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.

1. Introduction

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

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

- **Must:** This word indicates absolute technical requirement with which all projects must comply.
- **Should:** This word indicates that there may be valid reasons not to treat this point of guidance as an absolute requirement, but the full implications need to be understood and the case carefully weighed before it is disregarded. '**Should**' has been used in conjunction with technical standards that are likely to become widely implemented 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.

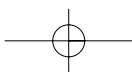
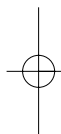
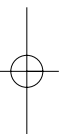
Those key words are used in the remainder of this document. When re-purposing this document in the context of the standards and guidelines for a specific programme, the authors **should** adapt the requirement levels specified in this document to those of their own contexts; authors **should** make appropriate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

Guidance

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

1.6 Summary

This document seeks to provide a core set of guidelines, rather than to attempt to reflect the different requirements of many different programmes and projects. Implementers will need to adapt these guidelines to the specific context in which they are operating, to select, to customize and to supplement as required. We hope that as a core, the MINERVA Technical Guidelines provide a useful starting point for many different contexts.



2. Projects and Planning

2.1 Introduction

A digitisation project has many dimensions and no two digitisation projects are identical. Each project varies according to the type of materials being digitised, the timescale, budget, staff skills and other factors. Some projects are completed in-house; some contract work externally while others involve collaboration between independent organizations. Each project will need to develop a project plan to fit its particular circumstances.

There are a number of formal methods for managing projects. It can be useful to follow one of these methods as this can help in transferring lessons-learned from one project to the next. Prince2 (Projects IN Controlled Environments) is an example of a formal methodology for managing projects, which is widely used in the UK and also internationally. Prince2 is a common sense approach that can be used on projects of all sizes; the full methodology does not need to be applied to all projects. Project managers **should** be aware that Prince2, or a similar methodology, should be applied selectively to meet the needs of their projects. More information about Prince2 is available at http://www.ogc.gov.uk/methods_prince_2.asp Available 2008-05-01.

A project can be defined as a temporary organization that is created to achieve a specific objective and then disbanded when the work has been completed (Prince2). It is triggered by a business need (for example the need to establish an online catalogue to support enquiry services) and takes place in the strategic context of an organisation. Projects may form part of a programme, but each project has:

- A life span and a defined life cycle (time)
- Defined products (specification)
- Activities to achieve the products
- A finite amount of resources (budget, equipment etc)
- An organizational structure with responsibility for managing and completing the project (people)

2. Projects and Planning

Projects may be said to have three primary objectives: **quality** (fitness for purpose or specification), **cost** (the budget) and **time** (to completion). Projects aim to achieve success in all these but project sponsors or managers may need to identify one of the objectives as being more important. For example, if the quality of the digital images is the most important objective it may be acceptable to take more time to complete the work or to spend additional budget on new equipment. **People** are fundamental to success in any project, providing good management, organization and motivation, and are central to balancing the project objectives (see figure 1).

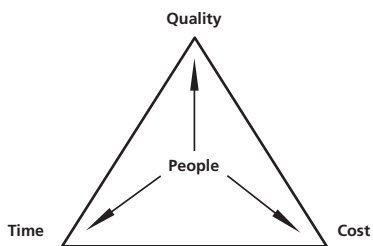


Figure 1: Triangle of Objectives

2.2 Project phases

Typically a project begins life following an initial idea and goes through a series of operational phases to its completion with the delivery of a new product (such as an online catalogue) and its evaluation. The project phases are:

Pre-project preparation

The aim of this phase is to make sure that everything that is needed to start the project is in place. It may include a feasibility study to evaluate options and recommend an approach. The aim of this phase is to establish:

- The **project design** which should cover why digitisation is beneficial, what will be digitised, how it will be digitised, the costs and resources required.
- How the project is to be funded.
- The senior management team (project board and project manager).
- Stakeholders and their quality expectations (user needs and requirements).

Getting started

The project begins only once funding is agreed and the go-ahead has been given by the project sponsors. The aim of this phase is to **initiate** the work of the project and to:

- **Plan** and cost the work.

2. Projects and Planning

- Refine or update the **business case**. This **should** include looking at long-term maintenance and use issues such as promotion, sustainability, administration and preservation.
- Define how the stakeholders' **needs and requirements** will be met (planning for quality).
- Assess any **risks** in carrying out the project and how they might be managed.

Implementing the project

In this phase the project's activities are carried out, project managers may establish:

- **Tasks** or work packages, allocating people, equipment, budget and time to each. Tasks may be contracted out to external suppliers or delivered in-house according to the project approach.
- **Evaluation** procedures and quality **benchmarks** – the process of obtaining feedback from users/stakeholders and measuring the quality of the products against their needs and requirements.
- **Monitoring** procedures – progress, risks, quality etc
- **Reporting** procedures – to keep sponsors, senior managers and project team members informed about the project's progress as appropriate.
- **Dissemination** - informing people outside the project about progress.

Most projects are divided into more than one **stage**. The sponsors normally review progress as each stage is completed making a **go-no go** decision before the project may continue to the next stage.

Closing a project

The final stage involves bringing the project to a formal close. Its objectives include:

- **Checking** that all products have been delivered and accepted and that any external. **Notifying** any suppliers of the close so they can plan for the return of resources or the submission of invoices.
- Identifying any **follow-on** actions, for example actions needed to maintain or support a new service.
- Planning for any post-project review.
- Producing a **final report** summarizing the projects activities and covering follow-on actions, lessons-learned.
- **Archiving** the project's records.

Guidance:

Framework for Building Good Digital Collections: Initiatives,
NISO with IMLS,
<<http://www.niso.org/publications/rp/framework3.pdf>>
Available 2008-09-01

2. Projects and Planning

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

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

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

Digitisation: a project planning checklist, AHDS
<http://ahds.ac.uk/creating/information-papers/checklist/index.htm>
 Available 2008-09-09

Managing Successful Programmes, OGC
http://www.ogc.gov.uk/guidance_managing_successful_projects.asp
 Available 2008-09-01

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

2.3 Planning

'A good collection-building initiative has a substantial design and planning component', NISO initiatives principle 1.

Planning is crucial to the success of any project or programme - even small projects need planning. Planning covers all aspects of the project from short and longer-term goals and objectives, constraints (time, costs, personnel, political factors etc.), selection, workflow, methodology, copyright issues, access, dissemination and evaluation.

A project manager is normally responsible for preparing the initial project plan during the initiation phase and then for updating the plan at the start of each project stage. Planning a stage involves looking at the details of the work required, identifying specific tasks and estimating the time and cost to complete and who will carry out the work. Updating the project plan at regular intervals is important - project plans are dynamic and must reflect the current situation. Most projects face "unexpected" challenges or events which may affect the timescales, costs and outcomes of the project. Updating the plan enables these unexpected events to be dealt with effectively.

2. Projects and Planning

There are a number of techniques that can help with project planning. For example, a **Work Breakdown Structure**, which describes the tasks and their relationships in a hierarchy. A **Gant chart** illustrates how tasks are scheduled within the project timeline. **Network analysis** can help to show the inter-dependencies between tasks. **PERT** (Programme Evaluation and Review Technique) is a technique for analyzing the tasks involved in a project and the time required to complete each task. Software applications are available to assist with these techniques. Which you use is likely to depend on the size of the project, the preferences of your organization or project sponsor and experience from previous projects.

Guidance

Project management software directory, Project Management Center
<http://www.infogoal.com/pmc/pmcsur.htm>
 Available 2008-09-09

2.4 People and Roles

"A good digital initiative has an appropriate level of staffing with necessary expertise to achieve its objectives." NISO initiatives, principle 2.

People have many different roles in the context of digitisation projects. All projects, at some point, are likely to need access to expertise in management and project management, budget and finance, legal issues, programming and systems administration, selecting content, preparing material for digitisation, digitisation, creating metadata and more. Some roles may be filled by the same person, while others require different people. Some people may work on the project full-time, while others for a few hours a month.

Projects may choose between three main strategies for accommodating the different roles and skills needed:

- in-house staffing,
- outsourcing or
- collaboration with one or more partner organizations.

Each strategy has advantages and drawbacks. Whichever is chosen, projects will need to make sure that staff involved have received adequate training in order for them to complete their tasks to an adequate level. Organisations should have appropriate training and development plans in place for staff.

Guidance

Handbook on Cost Reduction in Digitisation, section 3.3,
 MINERVA Plus, http://www.minervaeurope.org/publications/CostReductioninDigitisation_v1_0610.pdf
 Available 2008-05-01

2. Projects and Planning

Staff training, TASI

[<http://www.tasi.ac.uk/advice/managing/staff_training.html>](http://www.tasi.ac.uk/advice/managing/staff_training.html)

Available 2008-09-09

2.6 Managing Risks

Projects should carry out a risk assessment when developing their project plan. This allows the project manager to identify weak areas or potential problems – and to plan strategies for coping if the risk actually happens. Types of risk include:

- scheduling (not enough time allowed to complete a task)
- project scope creep (taking on new unplanned tasks)
- key skills in existing staff (lacking the skills to complete a task)
- time to appoint new staff
- staff leaving the project
- insufficient documentation of the workflow or methodology
- equipment failure
- insufficient IT support
- problems with commissioning work from external suppliers

Risk management involves identifying the possible risks, assessing their impact on the project (high, medium, low), their importance and developing plans to manage (or reduce) the impact of the risk should it occur. For example, the risk of staff lacking the key skills needed for the project can be managed through staff training.

Guidance:

Risk Assessment, TASI,

[<http://www.tasi.ac.uk/advice/managing/risk.html>](http://www.tasi.ac.uk/advice/managing/risk.html)

Available 2008-09-09

Risk Management and Contingency Planning, AHDS

[<http://ahds.ac.uk/creating/information-papers/risk-management/>](http://ahds.ac.uk/creating/information-papers/risk-management/)

Available 2008-09-09

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

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

In preparing 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. Projects **should** be aware of large-scale digitisation initiatives and methods for cost reduction such as outsourcing, automating digitization and metadata creation, streamlining workflow, continuous improvement and quality assurance. Projects **should** be aware of the NISO/IMLS Framework of Guidance for Building Good Digital Collections <<http://www.niso.org/publications/rp/framework3.pdf>> Available 2008-09-01.

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

FIAF (the International Federation of Film Archives) has developed guidelines which provide a code of ethics for staff involved in guardianship of "the world's moving image heritage" <<http://www.fiafnet.org/uk/members/ethics.cfm>> Available 2008-09-15. These guidelines may prove useful for those involved in the manipulation of other formats of cultural heritage resources.

3. Preparing for the digitisation process

3.1 Selecting Materials for Digitisation

"A good digital collection is created according to an explicit collection development policy that has been agreed upon and documented before building the collection begins." NISO collections principle 1.

Most digitisation projects involve the selection of material. Institutions generally have a particular set of material in mind when planning a project, selecting material to meet the needs of their target audiences and to reflect their overall collection development policy. Material may be chosen to meet the criteria of an external funding body, to enable collaboration with another institution or to coincide with a particular anniversary. Mass digitisation projects may involve little selection below the collection level (for example, digitizing an institution's collection of 19th century newspapers). At whatever level the selection is made, projects **should** reflect their selection policy in their business case and project plan (see section 2).

In the case of digitisation on demand, where institutions are offering end-users a service in which they create digital content on request, selection policy is less important.

However selected, a good digital collection consists of digital objects that have been developed according to a collection development policy, which aims to make sure that the content is created in line with best practice guidance and is described, actively managed, available, interoperable and sustainable over time.

Guidance:

Selection Procedures, TASI

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

Available 2008-09-10

Selection and Preparation of Materials, TASI

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

Available 2008-09-10

Framework Guidance for Building Good Digital Collections, NISO with IMLS,

<http://www.niso.org/publications/rp/framework3.pdf>

Available 2008-09-01

Creating and Managing Digital Content, CHIN

http://www.chin.gc.ca/English/Digital_Content/index.html

Available 2008-09-01

3. Preparing for the digitisation process

Handbook on Cost Reduction in Digitisation, Selection and Preparation for Digitisation, MINERVA,
http://www.minervaeurope.org/publications/CostReductioninDigitisation_v1_0b10.pdf
 Available 2008-05-01

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

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

3.2 Preparing Original Materials

Cataloguing originals

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

Cataloguing is important for:

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

Guidance:

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

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

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

3. Preparing for the digitisation process

Movement and Handling of Original Materials

The original materials may need to be cleaned or conserved before digitisation takes place. The time and cost of any such work should be taken into consideration in the project plan.

Different formats require different approaches to digitisation (for example, flat objects require a different approach to books or three-dimensional objects). There may be a need to plan for special handling, for example of fragile originals (medieval manuscripts), unsafe materials (nitrate film stock) or very large originals. Projects **should** consider the format of the original material when establishing their digital capture workflow.

Guidance:

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

Protecting the Physical Form, Joint RLG and NPO Preservation Conference, Guidelines for Digital Imaging,
<http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?fileid=0000073852:000007634487&reqid=18557>
 Available 2008-04-28

Selection and preparation of Materials, TASI,
<http://www.tasi.ac.uk/advice/creating/selection.html>
 Available 2008-09-10.

3.3 Staff Training

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

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

3.4 In-house Digitisation or Out-sourcing?

Establishing the appropriate environment in which digitisation will take place is important. It helps to make sure that the process is effective in creating usable digital resources and to keep any damaging effect on the original source materials to the minimum.

3. Preparing for the digitisation process

The digitisation may be carried out in-house by the institution or it may be outsourced to an external agency. Project **must** understand the factors involved in this choice, these include:

- The volume of similar original material to be digitised
- The fragility of the material and the risk of moving it outside the institution
- The organization of the physical collection and the associated catalogues
- The complexity of the digitisation process
- The availability of trained staff
- The availability of hardware and software

Reasons of outsourcing may include cost reduction (for large volumes), access to specialized equipment or practical limitations (space, people, equipment) within an institution. Reasons for carrying out the work in-house may include inability to move a collection, small volumes of material or the organization of the physical collection.

Guidance

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

Available 2008-04-28

Handbook on Cost Reduction in Digitisation, Outsourcing, MINERVA,
<http://www.minervaeurope.org/publications/CostReductioninDigitisation_v1_0610.pdf>

Available 2008-05-01

3.5 Hardware and Software

This document does not provide specific advice on the choice of digitisation hardware or software. Projects **must** demonstrate an awareness of the range that is available and are advised to consult the latest reviews and reports before purchasing equipment. The Canadian Heritage Information Network (CHIN) maintains a list of reviews that may be helpful: <http://www.chin.gc.ca/English/Digital_Content/Hardware_Software/index.html> Available 2008-09-01.

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.

Projects **must** ensure that the hardware and software selected has the functionality to produce digital objects of a quality that meets the

3. Preparing for the digitisation process

requirements of their expected uses, within acceptable constraints of cost. The hardware and software must be usable by the relevant project staff.

Hardware

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 computer.

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

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

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

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

Software

Projects **must** demonstrate an awareness of the use of software in image capture and image processing, and the hardware and software requirements of individual software products.

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

Open source software should be evaluated along with proprietary software packages. Consideration should be given to the European IDABC programme, national laws or directives promoting the adoption of open source software by public administrations (e.g. in countries including Italy, Germany, France and the UK). The potential benefits of open source software should be balanced against the potential risks, which may include the quality of documentation or the sustainability of the development community.

Guidance:

Image Capture: Hardware and Software, TASI
<<http://www.tasi.ac.uk/advice/creating/hwandsw.html>>
Available 2008-09-10

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

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

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

Reference

Hardware and Software Reviews, Canadian Heritage Information Network (CHIN)
http://www.chin.gc.ca/English/Digital_Content/Hardware_Software/index.html
 Available 2008-09-01.

3.6 The Digitisation Process

The Technical Advisory Service for Images (TASI) provides further guidance on the digitisation process. In addition a series of resources were developed by the former Arts and Humanities Data Service (AHDS) and these are still available.

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

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

Guidance:

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

Guides to Good Practice, AHDS,
<http://ahds.ac.uk/creating/guides/>
 Available 2008-08-28

3. Preparing for the digitisation process

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

Handbook on cost reduction in digitisation, Minerva,
<<http://www.minervaeurope.org/publications/costreduction.htm>>
Available 2008-08-28

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

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

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

4. Storage and Management of the Digital Master Material

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

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

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

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

Guidance:

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

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

4. Storage and Management of the Digital Master Material

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

4.1 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.) 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.

4.1.1 Text Capture and Storage

Character Encoding

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

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

Standards:

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

4. Storage and Management of the Digital Master Material

Extensible Markup Language (XML) 1.0

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

Available 2008-04-28

XHTML 1.0 The Extensible HyperText Markup Language

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

Available 2008-04-28

Guidance:

Jukka Korpela, A Tutorial on Character Code Issues

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

Available 2008-04-28

Character Encoding, Wikipedia,

http://en.wikipedia.org/wiki/Character_code

Available 2008-04-28

Document Formats

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

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

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

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

4. Storage and Management of the Digital Master Material

Standards:

ISO 8879:1986. Information Processing — Text and Office Systems — Standard Generalized Markup Language (SGML),
<<http://www.w3.org/MarkUp/SGML/>>
Available 2008-05-02

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

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

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

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

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

Other references:

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

Guidance:

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

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

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

4.1.2 Still Image Capture and Storage

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

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

Raster Images

When creating and storing raster images, two factors need to be considered these are 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:** high 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.

4. Storage and Management of the Digital Master Material

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

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

Standards:

Tagged Image File Format (TIFF)

<http://www.itu.int/itudoc/itu-t/com116/tiff-fx/docs/tiff6.pdf>

Available 2008-04-28

Joint Photographic Expert Group (JPEG)

<http://www.w3.org/Graphics/JPEG/>

Available 2008-04-28

JPEG Still Picture Interchange File Format (SPIFF)

<http://www.jpeg.org/public/spiff.pdf>

Available 2008-04-28

Guidance:

TASI: Advice: Creating Digital Images

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

Available 2008-04-28

JPEG, Wikipedia,

<http://en.wikipedia.org/wiki/JPEG>

Available 2008-04-28

Tagged Image File Format, Wikipedia,

http://en.wikipedia.org/wiki/Tagged_Image_File_Format

Available 2008-04-28

Graphic non-vector images

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

4. Storage and Management of the Digital Master Material

Standards:

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

Guidance:

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

Vector images

Vector images consist of multiple geometric objects (lines, ellipses, polygons, and other shapes) constructed through a sequence of commands or mathematical statements to plot lines and shapes. Vector graphics **should** be created and stored using an open format such as Scalable Vector Graphics (SVG), an XML language for describing such graphics. SVG drawings can be interactive and dynamic, and are scalable to different screen display and printer resolutions. Use of the proprietary Macromedia SWF format may also be appropriate (see 4.1.5 Multimedia below).

Standards:

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

Other references:

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

Guidance:

Vector graphics, Wikipedia
<http://en.wikipedia.org/wiki/Vector_graphics>
Available 2008-08-28

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

4. Storage and Management of the Digital Master Material

4.1.3 Video Capture and Storage

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

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

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

Standards:

The reference website for MPEG:

<<http://www.mpeg.org/>>

Available 2008-04-28

Guidance:

MPEG, Wikipedia

<<http://en.wikipedia.org/wiki/Mpeg>>

Available 2008-05-18

4.1.4 Audio Capture and Storage

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

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

Standards:

MP3, Fraunhofer IIS,

<<http://www.iis.fraunhofer.de/EN/bf/amm/projects/mp3/index.jsp>>

Available 2008-04-28

4. Storage and Management of the Digital Master Material

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

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

Guidance:

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

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

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

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

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

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

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

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

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

4. Storage and Management of the Digital Master Material

4.1.5 Multimedia

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

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

The Macromedia's proprietary SWF format (often referred to as Flash) may be a suitable medium for multimedia, however projects **should** explore a migration strategy so that they can move to more open formats once they become widely deployed. In addition, the use of text within the SWF format should be avoided, in order to support the development of multilingual versions.

The timed text (TT) authoring format is a content type that represents timed text media for the purpose of interchange among authoring systems. The Distribution Format Exchange Profile is a W3C candidate recommendation and **may** be an appropriate open standard for the exchange of timed text information (such as subtitles or captions).

Standards:

Synchronized Multimedia

<<http://www.w3.org/AudioVideo/>>

Available 2008-04-28

Macromedia Flash File Format (SWF) Specification License, Macromedia,

<<http://www.macromedia.com/software/flash/open/licensing/fileformat/>>

Available 2008-04-28

Timed Text (TT) Authoring Format 1.0 – Distribution Format Exchange Profile (DFXP), W3C Candidate Recommendation 16 November 2006,

<<http://www.w3.org/TR/ttaf1-dfxp/>>

Guidance:

SMIL, Wikipedia

<http://en.wikipedia.org/wiki/Synchronized_Multimedia_Integration_Language>

Available 2008-04-28

SWF, Wikipedia

<<http://en.wikipedia.org/wiki/SWF>>

Available 2008-04-28

4.1.6 GIS

GIS (Geographic Information Systems) can be used to integrate, store, edit, manage and present data which are spatially referenced (linked to location). The data that may be integrated in a GIS include raster images (e.g. digitised historic maps), vector images (e.g. maps captured using drawing software or data captured in the field using electronic measuring instruments), text and numeric data (e.g. databases describing the attributes of a location).

Geographic Information **should** be created and stored using non-proprietary and open data formats (such as the OpenGIS Geography Markup Language (GML)) and standards maintained by the Open Geospatial Consortium (OGC) and ISO; there are over 40 ISO standards which address a diverse range of functions. Use of proprietary data formats **may** be appropriate however projects **should** explore a migration strategy to open formats.

Standards:

ISO/TC 112

<<http://www.iso/112.org/>>

Available 2008-04-28

Open Geospatial Consortium standards and specifications

<<http://www.opengeospatial.org/standards>>

Available 2008-08-28

Guidance:

Geographic Information Systems, Wikipedia

<http://en.wikipedia.org/wiki/Geographic_information_system>

Available 2008-04-28

GIS File formats, Wikipedia

<http://en.wikipedia.org/wiki/Category:GIS_file_formats>

Available 2008-08-28

4.1.7 3D and Virtual Reality

3D models consist of a collection of points in three-dimensional space connected (by lines, triangles, curves or surfaces) to represent complex geometric objects. Models may represent either real-world objects, reconstructions based on the remains of a monument or building or virtual worlds created through the imagination of a designer.

3D scanners, electronic survey equipment or photogrammetry may be used to capture the geometry of real-world objects, buildings or monuments. The point data that is produced by these techniques is then

4. Storage and Management of the Digital Master Material

imported into 3D software as a 3D model. Data may be combined to represent complex shapes, such as the inside and outside of a building. CAD software is widely used for architectural objects and archaeological monuments, but a wide range of other 3D software packages is available. MeshLab is an open-source software programme for processing 3D meshes and data produced by 3D scanning.

3D software packages allow users to create and alter 3D models. For example, a reconstruction can be created by first importing point data captured from an archaeological monument and then using the software to add the missing portions (of walls or other elements). The software can then be used to create a realistic appearance by allowing surfaces to be rendered with images and lighting effects to be applied. 3D software is also used by designers, for example to produce models of new objects or to create virtual worlds.

Most 3D software packages enable models to be exported as a file for import into other applications (such as Virtual Reality). Animations and videos may be created.

Open standards are not well developed for 3D capture or 3D graphics. Laser scanners produce data in proprietary formats and the same is true for 3D software packages, including the open-source packages like Meshlab. Projects **may** choose to store CAD content using the proprietary dwg ("Drawing") format; this Autodesk format has become a defacto standard and is supported by other applications and also by the Open Design Alliance. Projects **may** choose to store CAD data using the proprietary dxf ("Drawing Exchange Format"), AutoCAD now publishes the specifications for the dxf versions used in its CAD software which supports the import of dxf files into other packages.

Virtual Reality integrates 3D models with text, sound and images to create computer-simulated environments in which users can interact (with a game, a virtual world or in some cases with each other). Virtual X3D is an ISO standard for virtual reality that has been developed by the Web 3D consortium from Virtual Reality Modelling Language (VRML) and which provides a system for the storage, retrieval and playback of real time graphics content embedded in applications. Virtual Reality models **should** be created using the X3D format.

COLLADA (COLLABorative Design Activity) defines an open standard XML schema for exchanging 3D files between 3D software applications such as Maya, 3ds Max, MeshLab, Blender and others. COLLADA is used as the native format by a number of applications, for example Google Earth. Projects **may** use COLLADA with X3D to develop 3D applications.

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

4. Storage and Management of the Digital Master Material

In some cases, it may be appropriate for cultural institutions to use managed 3-D virtual worlds (such as Second Life) to engage new audiences and display digital assets but it is not suggested that such environments are used for the creation or preservation of digital masters.

Standards:

Adobe Director and Shockwave

<http://www.adobe.com/uk/products/director/>

Available 2008-05-02

COLLADA, Khronos,

<http://www.khronos.org/collada/>

Available 2008-09-30

MeshLab

<http://meshlab.sourceforge.net/>

Available 2008-09-30

VRML

<http://www.web3d.org/x3d/specifications/vrml/>

ISO-IEC-14772-VRML97/

Available 2008-05-02

Virtual X3D

<http://www.web3d.org/about/overview/>

Available 2008-05-02

Guidance:

CAD: a Guide to Good Practice, ADS

<http://ads.ahds.ac.uk/project/goodguides/cad/>

Available 2008-05-02

Creating and Using Virtual Reality, VADS

http://vads.ahds.ac.uk/guides/vr_guide/

Available 2008-05-02

Web3D Consortium

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

Open Design Alliance

<http://www.opendesign.com/> Available 2008-05-02

Developing Web Applications with COLLADA and X3D, Khronos,

http://www.khronos.org/collada/presentations/Developing_Web_Applications_with_COLLADA_and_X3D.pdf

Available 2008-09-30

4. Storage and Management of the Digital Master Material

COLLADA, Wikipedia

<<http://en.wikipedia.org/wiki/COLLADA>>

Available 2008-09-30

DWG, Wikipedia

<<http://en.wikipedia.org/wiki/.dwg>>

Available 2008-05-08

DXF, Wikipedia

<http://en.wikipedia.org/wiki/AutoCAD_DXF>

Available 2008-09-30

X3D, Wikipedia

<<http://en.wikipedia.org/wiki/X3D>>

Available 2008-05-08

3D computer graphics software, Wikipedia

<http://en.wikipedia.org/wiki/3D_computer_graphics_software>

Available 2008-08-28

MeshLab, Wikipedia

<<http://en.wikipedia.org/wiki/MeshLab>>

Available 2008-09-30

4.2 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

4. Storage and Management of the Digital Master Material

storage medium. At least one copy **should** be kept at a location other than the primary site to ensure that they are safe in the case of any disaster affecting the main site. All transfers to portable media **should** be logged.

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

Guidance:

Preservation Management of Digital Materials

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

Available 2008-04-28

Using CD-R and DVD-R for Digital Preservation, TASI,

<http://www.tasi.ac.uk/advice/delivering/cdr-dvdr.html>

Available 2008-04-28

4.3 Preservation Strategies

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

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

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

The capture of metadata is a critical part of a migration-based preservation strategy (see 5.2.3). Metadata is required to support the management of the object and of the migration process, but furthermore, migration inevitably leads, at least in the longer term, to some changes in, or losses of, original

4. Storage and Management of the Digital Master Material

functionality. Where this is significant to the interpretation of the object, users will rely on metadata about the migration process- and about the original object and its transformations - to provide some understanding of the functionality provided in the original technological environment.

Guidance:

Preservation Management of Digital Materials Handbook

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

Available 2008-08-28

The State of Digital Preservation: An International Perspective

<<http://www.clir.org/PUBS/reports/pub107/contents.html>>

Available 2008-04-28

Planets - Preservation and Long-term Access through NETworked Services

<<http://www.planets-project.eu/>>

Available 2008-04-20

5. Metadata, Standards and Resource Discovery

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.

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

5.1 Metadata Standards

“Good metadata conforms to community standards in a way that is appropriate to the materials in the collection, users of the collection, and current and potential future uses of the collection” NISO Metadata Principle 1 and Principle 5 “Good metadata supports the long-term management, curation, and preservation of objects in collections”.

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),

5. Metadata, standards and resource discovery

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
<<http://www.collectionstrust.org.uk/specfaq>>
Available 2008-08-28

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

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

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

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

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

Guidance:

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

Framework of Guidance for Building Good Digital Collections,
Metadata principles, NISO with IMLS
<<http://www.niso.org/publications/rp/framework3.pdf>>
Available 2008-09-01

5.2.1 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 (for example a requirement to provide spatial coverage and temporal coverage separately), and **may** need to capture and store additional descriptive metadata to meet those requirements.

Standards:

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

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

5. Metadata, standards and resource discovery

Guidance:

Using Dublin Core

<<http://dublincore.org/documents/usageguide/>>

Available 2008-04-28

5.2.2 Administrative Metadata

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

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

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

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

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

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

Standards:

NISO Z39.87-2002 AIIM 20-2002 Data Dictionary — Technical Metadata for Digital Still Images

<http://www.niso.org/standards/resources/Z39_87_trial_use.pdf>

Available 2008-04-28

5.2.3 Preservation Metadata

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

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

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

Standards:

Reference Model for an Open Archival Information System (OAIS)
<<http://public.ccsds.org/publications/archive/b50x0b1.pdf>>
Available 2008-04-28

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

PREMIS data dictionary for preservation metadata, Version 2.0
March 2008
<<http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>>
Available 2008-09-24

PREMIS preservation metadata maintenance activity,
<<http://www.loc.gov/standards/premis/>>
Available 2008-09-24

PREMIS schema version 1.1,
<<http://www.loc.gov/standards/premis/schemas.html>>
Available 2008-09-24

Guidance:

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

5. Metadata, standards and resource discovery

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

5.2.4 Structural Metadata

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

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

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

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

Standards:

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

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

Guidance:

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

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

5.2.5 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 including scope, format, ownership, restrictions on access (NISO, Collections Principle 2). Description allows collections to be integrated into the wider body of existing digital collections and into digital services operating across these collections.

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

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

Standards:

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

References:

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

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

RSLP Collection Description
<<http://www.ukoln.ac.uk/metadata/rs1p/>>
Available 2008-04-28

5. Metadata, standards and resource discovery

Guidance:

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

The Institute of Museum and Library Services' Framework of Guidance for Building Good Digital Collections, Metadata Principles, NISO with IMLS
<http://www.niso.org/publications/rp/framework3.pdf>

5.3 Terms and Conditions of Use

"Good metadata includes a clear statement of the conditions and terms of use for the digital object" NISO Metadata principle 4.

Projects **should** clearly indicate under what terms and conditions or under which licence metadata and content can be re-used by third parties, such as Creative Commons licenses (see Section 9).

Standards:

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

Guidance:

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

5.4 Terminology Standards

"Good metadata uses authority control and content standards to describe objects and co-locate related objects" NISO Metadata principle 3.

Transmitting the information contained in metadata records requires more than a shared understanding of the metadata schema in use. It also depends on understanding the terms used as values in the metadata elements. This is normally achieved either by adoption of common terminologies or by adopting different terminologies where the relationships between terms in the different schemes are known.

Projects **should** use recognised multilingual terminological sources to provide values for metadata elements where possible. Only if no standard

terminology is available, local terminologies **may** be considered. Where local terminologies are deployed, information about the terminology and its constituent terms and their meaning **must** be made publicly available.

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

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

Reference:

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

The Institute of Museum and Library Services' Framework of Guidance for Building Good Digital Collections, Metadata Principles, NISO with IMLS
<<http://www.niso.org/publications/rp/framework3.pdf>>

5.4 Resource Discovery

"Good metadata supports interoperability" NISO Metadata Principle 2.

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

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

Projects **should** create one or more collection-level metadata records describing their collections as unit in an appropriate Collection Description Service. Projects **may** expose item-level metadata records describing individual digital resources within their collection(s).

Both collection-level and item-level metadata records **should** refer to 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

5. Metadata, standards and resource discovery

form of simple, unqualified Dublin Core metadata records and **may** also wish to provide item-level descriptions conforming to the MINERVA DC.Culture schema, the specification of metadata elements for Europeana or other specific application profiles.

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 (LOM).

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, DCMI

<<http://dublincore.org/documents/dces/>>

Available 2008-04-28

Dublin Core Collections Application Profile, DCMI

<<http://dublincore.org/groups/collections/collection-application-profile/>>

Available 2008-08-28

DC.Culture, MINERVA

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

Available 2008-04-28

IEEE Learning Object Metadata, IEEE

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

Available 2008-04-28

Guidance:

Dublin Core, Wikipedia,

<http://en.wikipedia.org/wiki/Dublin_core>

Available 2008-04-28

Provide Content, Europeana

<http://www.europeana.eu/provide_content.php>

Available 2008-08-29

Specification for the Metadata Elements for the Europeana Prototype, Europeana,

<http://www.europeana.eu/public_documents/Specification_for_metadata_elements_in_the_Europeana_prototype.pdf>

Available 2008-09-30

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

The Institute of Museum and Library Services' Framework of
 Guidance for Building Good Digital Collections, Metadata
 Principles, NISO with IMLS
[<http://www.niso.org/publications/rp/framework3.pdf>](http://www.niso.org/publications/rp/framework3.pdf)

5.5 Metadata, Ontologies and the Semantic Web

Sharing metadata records beyond the boundaries of a system requires the recipient to be able to interpret and process the data elements as the sender intends.

XML-based formats are widely employed for storing data and especially for **exchanging** data between programs, applications and systems. Conformance with a **community standard** XML Document Type Description (DTD) or XML Schema enables metadata records to be shared within that community. Where metadata is shared **beyond** the boundaries of a community (for example between museums, libraries and archives) **different** structural conventions may make the sharing of data more difficult and may require programming software to handle the different conventions.

Projects **may** wish to take advantage of the capacities to share and reuse data on the Web that are provided by a family of specifications coordinated by W3C's Semantic Web activity.

Within its framework of recommendations W3C's Semantic Web activity has specified several formal knowledge representation languages, like:

- Resource Description Framework (RDF, [<http://www.w3.org/RDF/>](http://www.w3.org/RDF/)),
- Simple Knowledge Organisation System (SKOS, [<http://www.w3.org/2004/02/skos/>](http://www.w3.org/2004/02/skos/)),
- Web Ontology Language (OWL, [<http://www.w3.org/TR/owl-features/>](http://www.w3.org/TR/owl-features/)) etc.

Each language offers a different representational capacity with reference to its formal syntax and semantics and is appropriate for different types of information exchange depending on the complexity.

RDF is considered as a general framework for describing web resources in the triple form, subject-predicate-object. Subject denotes a resources, predicate a specific aspect (e.g. a property) of the resource and object a relationship of the resource with a specification (e.g. the value of the property). Based on a simple data model, RDF is appropriate for providing the simple semantics of web resources and metadata records in

5. Metadata, standards and resource discovery

a formal machine-understandable way. But the simple syntax restricts its appropriateness for representing complex information structures that need more syntactic and semantic expressivity. For example, RDF has no internal way to distinguish between different types of resources (like schemas, data, concepts, roles, properties, hierarchies or taxonomies etc).

The more expressive **Web Ontology Language (OWL)** was specified by W3C to provide greater machine interpretability of web resources. The expressive nature of OWL empowers web agents to try to build flexible information sharing and reuse infrastructures. However the expressivity and interpretational power of OWL is not always necessary. In some digital library applications it is sufficient to represent simple taxonomies, thesauri and classifications semantically. In such cases, simplicity can be more important than expressivity, even if the inferential power is reduced. W3C was motivated for similar reasons to standardize a family of formal languages appropriate for the representation of controlled and structured vocabularies.

Simple Knowledge Organisation System (SKOS) is a recommendation of W3C for the representation of thesaurus taxonomies. The SKOS standard is built on top of RDF language and can be used to facilitate the semantic retrieval of metadata and for thesaurus alignment.

SPARQL is an RDF query language (the name stands for *Simple Protocol and RDF Query Language*) is a recommendation of W3C. SPARQL allows for a query on triple patterns.

Standards:

W3C Semantic Web Activity

<http://www.w3.org/2001/sw/>

Available 2008-04-28

Web Ontology Language (OWL), W3C,

<http://www.w3.org/2001/sw/WebOnt/>

Available 2008-04-28

OWL Web Ontology Language Reference, W3C, 10 Feb 2004,

<http://www.w3.org/TR/owl-ref/>

Available 2008-04-28

CIDOC Conceptual Reference Model (CRM)

<http://cidoc.ics.forth.gr/>

Available 2008-04-28

SPARQL Query Language for RDF, W3C,

<http://www.w3.org/TR/rdf-sparql-query/>

Available 2008-09-23

Guidance:

RDF Primer

<<http://www.w3.org/TR/rdf-primer/>>

Available 2008-04-28

OWL Web Ontology Language Overview

<<http://www.w3.org/TR/owl-features/>>

Available 2008-04-28

The ABC Ontology and Model

<http://metadata.net/harmony/JODI_Final.pdf>

Available 2008-04-28

Semantic Web, Wikipedia

<http://en.wikipedia.org/wiki/Semantic_web>

Available 2008-04-28

Web Ontology Language, Wikipedia

<http://en.wikipedia.org/wiki/Web_Ontology_Language>

Available 2008-04-28

SPARQL, Wikipedia

<<http://en.wikipedia.org/wiki/SPARQL>>

Available 2008-09-23

Metadata Sharing and XML, UKOLN

<<http://www.ukoln.ac.uk/interop-focus/gpg/Metadata/>>

Available 2008-09-23

5.6 OAI-PMH and Metadata Harvesting

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

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

Standards:

Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)

<<http://www.openarchives.org/>>

Available 2008-04-28

5. Metadata, standards and resource discovery

Guidance:

OAI FAQs

<<http://www.openarchives.org/documents/FAQ.html>>

Available 2008-04-28

Open Archives Initiative, Wikipedia,

<http://en.wikipedia.org/wiki/Open_Archives_Initiative>

Available 2008-04-28

5.7 Distributed Searching

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

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

Standards:

SRW: Search/Retrieve Web Service

<<http://www.loc.gov/standards/sru/>>

Available 2008-04-28

Z39.50 Maintenance Agency

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

Available 2008-04-28

Bath Profile

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

Available 2008-04-28

Guidance:

Z39.50, Wikipedia,

<<http://en.wikipedia.org/wiki/Z39.50>>

Available 2008-04-28

What is SRW/U?, Techessence,

<<http://techessence.info/node/48>>

Available 2008-04-28

5.8 Syndication / Alerting

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

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

Standards:

RDF Site Summary (RSS) 1.0

[<http://purl.org/rss/1.0/spec>](http://purl.org/rss/1.0/spec)

Available 2008-04-28

RSS 2.0

[<http://blogs.law.harvard.edu/tech/rss>](http://blogs.law.harvard.edu/tech/rss)

Available 2008-04-28

The Atom Syndication Format

[<http://atompub.org/2005/04/18/draft-ietf-atompub-format-08.html>](http://atompub.org/2005/04/18/draft-ietf-atompub-format-08.html)

Available 2008-04-28

Guidance:

RSS, Wikipedia

[<http://en.wikipedia.org/wiki/Rss>](http://en.wikipedia.org/wiki/Rss)

Available 2008-04-28

Podcast, Wikipedia

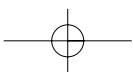
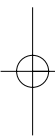
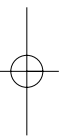
[<http://en.wikipedia.org/wiki/Podcast>](http://en.wikipedia.org/wiki/Podcast)

Available 2008-04-28

Video podcast, Wikipedia

[<http://en.wikipedia.org/wiki/Vodcast>](http://en.wikipedia.org/wiki/Vodcast)

Available 2008-04-28



6. Publishing on the Web

Projects **should** be aware of the ten MINERVA quality principles which state that a good quality cultural website must:

- be transparent, clearly stating the identity and purpose of the website, as well as the organisation responsible for its management
- select, digitise, author, present and validate content to create an effective website for users
- implement quality of service policy guidelines to ensure that the website is maintained and updated at an appropriate level
- be accessible to all users, irrespective of the technology they use or their disabilities, including navigation, content, and interactive elements
- be user-centred, taking into account the needs of users, ensuring relevance and ease of use through responding to evaluation and feedback
- be responsive, enabling users to contact the site and receive an appropriate reply. Where appropriate, encourage questions, information sharing and discussions with and between users
- be aware of the importance of multi-linguality by providing a minimum level of access in more than one language
- be committed to being interoperable within cultural networks to enable users to easily locate the content and services that meet their needs
- be managed to respect legal issues such as IPR and privacy and clearly state the terms and conditions on which the website and its contents may be used
- adopt strategies and standards to ensure that the website and its content can be preserved for the long-term

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

Guidance:

MINERVA Quality Principles for Cultural web sites, MINERVA
<<http://www.minervaeurope.org/qau/qualityprinciples.htm>>
Available 2008-08-29

6. Publishing on the Web

6.2 Browsers and Protocols

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

Standards:

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

Guidance:

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

6.3 Accessibility

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

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

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

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

Standards:

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

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

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

Guidance:

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

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

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

6.4 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.

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

6. Publishing on the Web

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

Standards:

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

Guidance:

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

6.5 Authenticity

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

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

Guidance:

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

6.6 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. This may include OpenID or SAML-based infrastructures, such as Shibboleth, which is increasingly being deployed in formal education contexts.

Standards:

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

Security Assertion Markup Language
 <<http://saml.xml.org/saml-specifications>>
 Available 2008-09-28

Guidance:

OpenID, Wikipedia
 <<http://en.wikipedia.org/wiki/OpenID>>
 Available 2008-09-28

Security Assertion Markup Language, Wikipedia
 <<http://en.wikipedia.org/wiki/SAML>>
 Available 2008-09-28

Shibboleth, Wikipedia
 <[http://en.wikipedia.org/wiki/Shibboleth_\(Internet2\)](http://en.wikipedia.org/wiki/Shibboleth_(Internet2))>
 Available 2008-09-28

6.7 Search Engine Optimisation

Projects **should** take steps to improve the visibility and ranking of their Web site by search engines. Search engine optimization includes:

- Following best practices in web site design by separating style from content, minimizing JavaScript and streamlining code to allow search engines to crawl, index and rank web pages more easily.
- Using clear and simple language appropriate for the site's content. Making sure that the free text on the website includes the words that audiences are likely to use when searching on search engines.

6. Publishing on the Web

- Providing a text equivalent for all non-text elements (such as still and moving images) in an alt or longdesc tag to allow search engines to understand and index the content.
- Providing text links give search engines important additional information about the content of the target page. Including redundant text links for image maps such as menus.
- Specifying the language of all documents, to enable correct indexing by search engines.
- Making sure that web-pages are usable when scripts, applets or other programmatic objects are turned off. Search engines do not read scripts.
- Registering your web-site with directories.
- Building links. Sending your website URL with a brief description to the webmasters of other cultural organizations, tourism authorities etc. Search engines take into consideration how many other Web sites link to your site when determining its ranking in search results.

Guidance

Search Engine Optimization, Wikipedia

[<http://en.wikipedia.org/wiki/Search_engine_optimization>](http://en.wikipedia.org/wiki/Search_engine_optimization)

Available 2008-09-23

Hagans, Andy, 'High Accessibility is effective search engine optimization', A List Apart,

[<http://www.alistapart.com/articles/accessibilityseo>](http://www.alistapart.com/articles/accessibilityseo)

Available 2008-09-23

'Producing Online Heritage Projects: Getting Ready to Launch: Search Engine Registration', CHIN

[<http://www.chin.gc.ca/English/Digital_Content/Producing_Heritage/searchengine.html>](http://www.chin.gc.ca/English/Digital_Content/Producing_Heritage/searchengine.html)

Available 2008-09-23

6.8 Current and Future Trends: Web 2.0-3.0

In recent years, technologies have emerged which are transforming the way that the Web is used and supporting creativity, information sharing, collaboration and new functionality for users. Web 2.0 saw the emergence of web-based communities and hosted services, such as social-networking sites, image sharing, wikis, blogs and folksonomies. Web 3.0 anticipates the emergence of the Semantic web.

The popularity of these new web-services and of social networking is generating much interest from the cultural heritage sector.

6.8.1 Technologies

The technologies are constantly evolving but include:

- Protocols for **syndicating content** including RSS, RDF and Atom
- **Web services and Web APIs** allowing machine-to-machine access to data and functions such as REST and SOAP
- Publishing software to enable the creation of **blogs**
- **Wiki** software to support the creation of web pages by a community of users
- Tools for **social tagging** and for producing folksonomies
- Web development techniques (such as **Ajax**) that support the creation of interactive applications in agile development environments
- Platforms to support multi-user environments
- And more

Web 2.0 saw the emergence of commercially hosted services (such as those offered by companies such as Google and Yahoo) which can be used in **mashups** to integrate external and in-house content and to deliver or augment services.

Projects **should** demonstrate awareness of the Web Services family of specifications, especially REST and SOAP version 1.2 and also the Web Services Description Language (WSDL). Projects **should** be aware of the potential for mashups, where web services (or APIs) are used to combine content from more than one source to develop new services, such as web-based mapping services.

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

Standards:

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

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

Guidance:

Web 2.0, Wikipedia
<http://en.wikipedia.org/wiki/Web_2.0>
Available 2008-09-01

Web 3.0, Wikipedia
<http://en.wikipedia.org/wiki/Web_3.0>
Available 2008-09-01

6. Publishing on the Web

Mashup, Wikipedia

<[http://en.wikipedia.org/wiki/Mashup_\(web_application_hybrid\)](http://en.wikipedia.org/wiki/Mashup_(web_application_hybrid))>

Available 2008-08-29

SOAP Version 1.2 Part 0: Primer

<<http://www.w3.org/TR/soap12-part0/>>

Available 2008-04-28

Web service, Wikipedia

<http://en.wikipedia.org/wiki/Web_services>

Available 2008-04-28

Universal Description Discovery and Integration, Wikipedia

<<http://en.wikipedia.org/wiki/UDDI>>

Available 2008-04-28

Handbook on cultural web user interaction, MinervaEC

<<http://www.minervaeurope.org/publications/handbookwebusers.htm>>

Available 2008-09-09

6.8.2 User perspectives

A discussion of current and future trends in Web developments from a user perspective can be found in the MINERVA **Handbook on cultural web user interaction**: <<http://www.minervaeurope.org/publications/handbookwebusers.htm>> Available 2008-09-09.

The Handbook deals with the relationship between user and web application in the light of the developments and new online models that have emerged in recent years. This is a practical manual focusing on interaction with users on the web that also investigates the current Internet trends, strongly orientated towards collaborative functionality, user interaction, social networks and sharing, the evolution of the Web 2.0 and the new challenges of the Web 3.0.

It is worth noting that these new technologies can offer alternative approaches to Web accessibility to the recommendations described in WAI's WCAG 1.0 guidelines. For example JavaScript and Ajax can be used to provide usable and accessible environments.

6.8.3 Factors to Take into Consideration

There are **risks** associated with the use of a Web 2.0-3.0 approach to the development and deployment of services. Some of the risks are given below, together with approaches to assessing and managing them:

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

Table 1: Risk assessment and management approaches

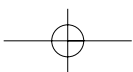
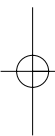
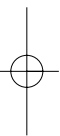
The possible disadvantages of using Commercial web services include:

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

Projects **should** also assess the risks of not using Web 2.0 services, such as missed opportunity costs, the cost of in-house development, the risk of losing staff and failing to meet user expectations etc.

The **benefits** of a Web 2.0-3.0 approach include:

- Delivery of services that are attractive to individuals for both leisure and professional use, and also to institutions who may also make use of these platforms.
- costs measured in staff hours of content management, rather than technological development
- intuitive and free upload of content (images, videos, dynamic maps, etc.), and
- seamless dissemination of content across social networks.



7. Delivery formats

It is expected that end-user access to resources will be primarily through the use of Internet protocols. Preparation for delivery 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.

Video and audio may be made available either for download or for streaming. With streaming, instead of 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 is 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 4.1).

7.1 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 **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, organisation structure of the provider service or any other information that is likely to change within the lifetime of the resource.

7. Delivery formats

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

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

Standards:

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

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

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

Guidance:

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

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

Persistent Identifiers, PADI
<<http://www.nla.gov.au/padi/topics/36.html>>
Available 2008-08-29

7.2 Delivery of Text

Character Encoding

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

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

Guidance:

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

Document Formats

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

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

It should also be noted that the Open XML (OOXML) and the OpenDocument 1.2 (ODF) formats are in the process of being standardised as open standards.

Standards:

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

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

Information technology — Open Document Format for Office Applications (OpenDocument) v1.0
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43485
 Available 2008-08-29

PDF 1.7, ISO 32000-1:2008, ISO
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=51502
 Available 2008-08-29

PDF 1.4, ISO 19005-1:2005, ISO
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=38920
 Available 2008-08-29

Reference:

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

7. Delivery formats

Guidance:

OpenDocument, Wikipedia,
<<http://en.wikipedia.org/wiki/OpenDocument>>
Available 2008-08-29

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

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

7.3 Delivery of Still Images

Photographic images

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

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

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

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

Graphic non-vector images

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

Graphic vector images

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

7.4 Delivery of Video

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

Downloading

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

Streaming

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

Guidance:

The Reference Website for MPEG

<<http://www.mpeg.org/>>

Available 2008-08-29

ASF, Wikipedia

<http://en.wikipedia.org/wiki/Advanced_Streaming_Format>

Available 2008-08-29

AVI, Wikipedia

<<http://en.wikipedia.org/wiki/AVI>>

Available 2008-08-29

QuickTime, Wikipedia

<http://en.wikipedia.org/wiki/Apple_Quicktime>

Available 2008-08-29

WMV, Wikipedia

<http://en.wikipedia.org/wiki/Windows_Media_Video>

Available 2008-08-29

7.5 Delivery of Audio

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

Downloading

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

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

Streaming

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

7. Delivery formats

Guidance:

AIFF, Wikipedia

<<http://en.wikipedia.org/wiki/AIFF>>

Available 2008-08-29

AU, Wikipedia

<http://en.wikipedia.org/wiki/Au_file_format>

Available 2008-08-29

MP3, Wikipedia,

<<http://en.wikipedia.org/wiki/MP3>>

Available 2008-04-28

RealAudio, Wikipedia

<<http://en.wikipedia.org/wiki/RealAudio>>

Available 2008-08-29

WAV, Wikipedia

<<http://en.wikipedia.org/wiki/WAV>>

Available 2008-04-29

WMA, Wikipedia

<http://en.wikipedia.org/wiki/Wma_audio_file>

Available 2008-08-29

Audio streaming, Wikipedia,

<http://en.wikipedia.org/wiki/Audio_streaming>

Available 2008-04-28

7.6 Delivery of Virtual Reality

Projects making use of three-dimensional virtual reality **must** consider the needs of users accessing their site using typical computers and Internet 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.

Projects **must** be aware that in order to view virtual reality models online users will need to install a plug in (such as the Cortona3D or SwirlX3D) for their web browser.

7. Delivery formats

Projects **must** be aware that the majority of their users are likely to have lower bandwidth and computers with lower specifications than those used to generate or test the models. Projects **must** therefore consider the usability of their models in such conditions, and **must** test them using typical connections and on home, school, or library computer systems with a variety of typical operating systems, browsers and standard plug-ins.

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

Projects **may** wish to consider offering alternative forms of access. For example, projects may consider offering users a video of a fly-through of a Virtual Reality Model or enabling users to move through a series of panoramas based on images of a Virtual world using Apple's QuickTime VR (QTVR) format.

Projects **may** wish to consider using managed 3-D virtual worlds (such as Second Life) to engage new audiences and display digital assets.

Standards:

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

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

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

7.7 Delivering Geographic Information

Much cultural content has a grounding in place, and this offers one powerful means by which content might be grouped or retrieved. Geographic Information Systems (GIS) are software applications specifically designed to store, manipulate and retrieve place-based information, which are widely deployed within the historic environment sector. Much place-based information is also stored within traditional databases, for example the birth-place of a person associated with a museum object.

It is not necessary to install and maintain a GIS to make simple images of location maps available on a website. Maps may also be generated using Application Programming Interfaces (APIs) or web services (see 8.4 below) to bring together cartographic data and other place-based information from distributed sources to produce maps. An example is the use of map data from Google Maps via Google's free web mapping API to add a map layer to cultural heritage (or other) information using place-names or spatial references within the dataset.

7. Delivery formats

Projects **should** be aware of the potential to develop web mapping services and **must** ensure that they can support GI implementation in the future. Mapping APIs or the OAI-PMH harvesting of metadata (see Section 5.6) can then allow the presentation of data through external systems.

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

- Projects seeking to employ a GIS **must** obtain appropriate permissions for use of any map data from third parties, ensuring that licences extend to delivering services to their defined audiences via their selected delivery channels.
- Projects **must** ensure that data sets combined for the purposes of delivering their service are of similar scale and resolution, and appropriate for use together in this manner.
- Commercial GIS products selected for use **should** comply with emerging industry standards from the Open GIS Consortium.
- Projects **must** make use of and declare use of an appropriate standard co-ordinate reference system when recording spatial data.
- Projects **must** make use of and declare use of appropriate national standards for the recording of street addresses.

Standards:

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

Guidance:

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

Google Maps, Wikipedia
<http://en.wikipedia.org/wiki/Google_Maps>
Available 2008-08-29

Mashup, Wikipedia
<[http://en.wikipedia.org/wiki/Mashup_\(web_application_hybrid\)](http://en.wikipedia.org/wiki/Mashup_(web_application_hybrid))>
Available 2008-08-29

Web mapping, Wikipedia
<http://en.wikipedia.org/wiki/Web_mapping>
Available 2008-08-29

8. Reuse and Re-purposing

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

8.1 Learning Resource Creation and Re-use

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

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

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

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

Standards:

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

8. Reuse and Re-purposing

IMS Global Learning Consortium, Inc.

[<http://www.imsproject.org/>](http://www.imsproject.org/)

Available 2008-04-28

IMS Content Packaging.

[<http://www.imsproject.org/content/packaging/>](http://www.imsproject.org/content/packaging/)

Available 2007-10-30

Guidance:

Learning object metadata, Wikipedia

[<http://en.wikipedia.org/wiki/Learning_object_metadata>](http://en.wikipedia.org/wiki/Learning_object_metadata)

Available 2008-04-28

Insight: knowledge base for new technology and education,
European Schoolnet

[<http://insight.eun.org/ww/en/pub/insight/index.htm>](http://insight.eun.org/ww/en/pub/insight/index.htm)

Available 2008-08-29

9. Intellectual Property Rights, Copyright, Licencing and Sustainability

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.** Much published material (including text, images, music, sound recordings and moving images) is likely to be in copyright, as copyright generally lasts for 70 years from the death of the creator of the work. Written permission from the copyright owner **must** be obtained before any digitisation commences.
- **Orphan Works.** The use of copyright works without permission (beyond a number of exceptions within the legislation) is an infringement of moral and economic rights. One of the major stumbling-blocks in mass digitisation projects is that many of the works to be digitised are still protected by copyright, and that gaining the vast range of permissions required entails a huge logistical and administrative overhead. In many cases, the current owners of copyright are unknown or untraceable, creating a class of works which has come to be known as 'orphan works'. The only currently available strategy is to adopt standardized due diligence processes involving protocols for researching contacts for rights-holders, seeking permissions, duly recording the process at each stage, so that diligence and good faith can be demonstrated. Although this offers no protection under the law, there is a weight of expectation that it would be taken into account by parties considering litigation, not least in terms of the likely scale of any resulting settlement. Orphan works **should** only be digitised once a due diligence process has been undertaken and documented.

9. Intellectual Property Rights, Copyright, Licencing and Sustainability

- **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.

9.1 Identifying, Recording and Managing Intellectual Property Rights

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

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

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

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

Guidance:

Creating Digital Resources for the Visual Arts: Standards and Good Practice

<http://vads.ahds.ac.uk/guides/creating_guide/contents.html>

Available 2008-04-28

European Digital Library High Level Group Sector-specific guidelines on diligence search criteria for orphan works

<http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/orphan/guidelines.pdf>

Available 2008-08-28

Intellectual property Rights Overview

<<http://www.jisclegal.ac.uk/ipr/IntellectualProperty.htm>>

Available 2008-04-28

9. Intellectual Property Rights, Copyright, Licencing and Sustainability

MINERVAeC IPR Guide

<http://www.minervaeurope.org/IPR/IPR_guide.html>

Available 2008-08-28

UK Intellectual Property Office

<<http://www.ipo.gov.uk/home.htm>>

Available 2008-04-28

World Intellectual Property Organization

<<http://www.wipo.org/>>

Available 2008-04-28

9.2 Safeguarding Intellectual Property Rights

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

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

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

The type of information required includes:

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

9.2.1 Creative Commons

The Creative Commons organisation has developed a set of licences which are designed to enable rights-owners to allow the re-use of material under conditions chosen by the rights-owner, typically allowing re-use in educational or not-for-profit contexts.

Projects that own the rights in the content that they are digitising **may** wish to assign a Creative Commons licence to their resources.

9. Intellectual Property Rights, Copyright, Licencing and Sustainability

Standards:

Creative Commons

<<http://www.creativecommons.org/>> Available 2008-04-28**Guidance:**

Creative Commons, Wikipedia

<http://en.wikipedia.org/wiki/Creative_Commons>

Available 2008-04-28

9.2.2 Planning for sustainability

Projects must plan to ensure sustainability of the service and of content, and should develop a business model for the service that is not reliant upon external funding. The licencing of the content, perhaps with different licences for different user communities, may also include the adoption of licences designed to generate revenue streams.

Guidance:

Ithaka, Sustainability and Revenue Models for Online Academic Resources

<www.jisc.ac.uk/media/documents/themes/eresources/sca_ithaka_sustainability_report-final.pdf>

Available 2008-09-02

9.2.3 Watermarking and Fingerprinting

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

Watermarking is the embedding of a permanent mark within a file that can subsequently be used to prove image origination or image copyright. This is normally achieved by integrating the watermark with the image data in such a way that it is virtually impossible to remove. Watermarks can be visible, invisible or a combination of both. In all cases the watermark is introduced in such a way that there is minimum distortion of the original image. Invisible watermarks **must** be able to withstand the image being cropped, rotated, compressed or transformed. As well as watermarking images before they are distributed, images can be fingerprinted dynamically at delivery time i.e. as the image is downloaded from a Web site. When this is done, other information such as username, date, time, IP address etc. can be encoded as part of the watermark. This makes each instance of download unique and traceable through a transaction database enabling tracking of who is downloading images. Similar techniques can be used in audio and video media.

Guidance:

Purloining and Pilfering, Web Developers Virtual Library

<<http://www.wdvl.com/Authoring/Graphics/Theft/>>

Available 2008-04-28

APPENDIX 1.

About this Document

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

Maintenance

These Guidelines were developed by the MINERVA project and all comments and suggestions for changes and updates should be submitted to the project. For more information about the MINERVA project see: <http://www.minervaeurope.org/> available 2008-09-20.

Links to resources

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

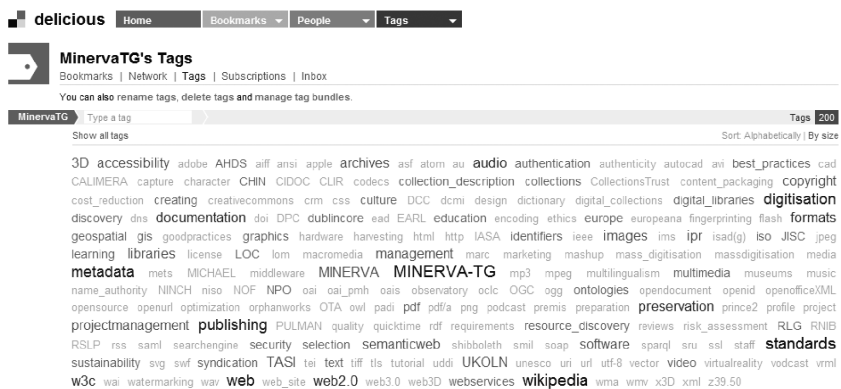


Figure 2: The tag cloud produced by the MINERVA technical guidelines' bookmarks on Delicious.

APPENDIX 1. About This Document

Links to guidance in Wikipedia

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

APPENDIX 2.

Glossary

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

AHDS	Arts and Humanities Data Service.
API	Application Programming Interface.
ASF	Advanced Systems Format.
AVI	Audio Video Interleave.
CDWA	Categories for the Description of Works of Art.
CHIN	Canadian Heritage Information Network.
CSS	Cascading Style Sheets
dpi	Dots per inch.
DCC	Digital Curation Centre.
DCMES	Dublin Core Metadata Element Set.
DNS	Domain Names System.
DOM	Document Object Model.
DTD	Document Type Definition.
DWG	Drawing.
EAD	Encoded Archival Description.
EMII	European Museums' Information Institute.
HEDS	Higher Education Digitisation Service
FIAF	International Federation of Film Archives.
GIF	Graphical Interchange Format.
GIS	Geographic Information System
GML	Geography Markup Language.
HTML	Hypertext markup Language.
HTTP	HyperText Transfer Protocol
IDABC	Interoperable Delivery of European eGovernment Services to public Administrations, Businesses and Citizens.
IETF	Internet Engineering Task Force.
ISAAR(CPF)	International Standard Archival Authority Record for Corporate Bodies
ISAD(G)	General International Standard Archival Description
ISO	International Standards Organisation.
IMLS	Institute of Museum and Library Services.
IMS	IMS Global Learning Consortium.
JISC	Joint Information Systems Committee.

APPENDIX 2. Glossary

JPEG	Joint Photographic Expert Group.
JPEG/SPiFF	JPEG Still Picture Interchange File Format.
LOM	Learning Object Metadata
MARC	MAchine Readable Cataloguing
METS	Metadata Encoding and Transmission Standard.
MPEG	Moving Picture Experts Group.
NISO	National Information Standards Organization.
NOF	New Opportunities Fund.
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OAIS	Open Archival Information System.
OCLC	Online Computer Library Center.
OGC	Open Geospatial Consortium.
OWL	Web Ontology Language.
PDF	Portable Document Format.
PLANETS	Planets - Preservation and Long-term Access through NETworked Services
PNG	Portable Network Graphics.
PRINCE2	PRojects IN Controlled Environments – a project management methodology.
RDF	Resource Description Framework.
RFC	Request For Comments.
RfP	Request for Proposal.
RLG	Research Libraries Group.
RNIB	Royal National Institute for the Blind.
RSS	Really Simple Syndication
SGML	Standard Generalized Markup Language.
SKOS	Simple Knowledge Organisation System.
SOAP	Simple Object Access Protocol.
SRW	Search/Retrieve Web Service
SSL	Secure Sockets Layer.
SVG	Scalable Vector Graphics.
SWF	Shockwave Flash.
TASI	Technical Advisory Service for Images.
TCP/IP	Transmission Control Protocol/Internet Protocol.
TEI	Text Encoding Initiative.
TIFF	Tagged Image File Format.
UDDI	Universal Description, Discovery and Integration
URI	Uniform Resource Identifier
VRML	Virtual Reality Modelling Language.
WAI	Web Accessibility Initiative.
WCAG	Web Content Accessibility Guidelines.
WMA	Windows Media Audio.
WMV	Windows Media Video
WSDL	Web Services Description Language
XHTML	eXtensible HyperText Markup Language.
XML	eXtensible Markup Language.

APPENDIX 3. Business Models: Web 2.0-3.0

Web 2.0-3.0 approaches are reshaping some of the traditional business models which underpin the delivery of services to the public. A primary characteristic of these approaches is the principle that *use has value*. Traditional transactional models tend to depend on a direct relation between service and payment (whether this payment is by the user or front-loaded in the form of investment or subsidy).

The new approaches tend to favour the downstream revenue model. In this case, the cost of developing and delivering a service is recouped not from the user, but usually from a 3rd party whose business depends on highly focused or targeted access to particular user groups.

A popular alternative is also to consider the longer-term reputational value which arises from the delivery of the service. In many cases, the initial cost of development can be offset against the equivalent value of establishing an ongoing relationship with a target user group (and hence can be regarded as an investment in marketing and brand awareness).

These models hold a particular appeal in a public service environment, which tends to require free-at-point-of-use access to content and services. It also offers a potential route to enhancing the sustainability of a given service by creating a new form of value which can be transacted with 3rd parties (reducing the dependence on ongoing grant income).

Finally, the participative nature of Web 2.0-3.0 approaches offers a potential route to reducing the medium-term costs of service development. If the capacity to deliver and maintain a service is localized within an organization, then the organization must resource 100% of the overhead involved.

If, however, the user is given the opportunity to participate in the development of services, this decentralizes the resourcing of the service, and hence reduces the overhead cost to the organization itself. Models such as crowdsourcing (opening up development to a distributed user community) can provide the organization with significantly increased capacity which would otherwise be beyond its means.

APPENDIX 3. Business models: Web 2.0-3.0

References

Everall, I and Ormes, S., 'Charging and Networked Services' EARL, Library Association and UKOLN,
<<http://www.ukoln.ac.uk/public/earl/issuepapers/charging.html>>
Available 2008-09-16

Tanner, S. and Deegan, M., 'Exploring Charging Models for Digital Library Cultural Heritage', Ariadne Issue 34, 2003,
<<http://www.ariadne.ac.uk/issue34/tanner/>>
Available 2008-09-16

CALIMERA, 'Business issues and Business models for the delivery of technology solutions in local cultural institutions',
<<http://www.calimera.org/Lists/Resources%20Library/Technologies%20and%20research%20for%20local%20cultural%20services/Business%20Issues%20and%20Models%20for%20Local%20Cultural%20Institutions.pdf>>
Available 2008-09-16

Crowdsourcing, Wikipedia
<<http://en.wikipedia.org/wiki/Crowdsourcing>>
Available 2008-09-16

Mass Collaboration, Wikipedia
<http://en.wikipedia.org/wiki/Mass_Collaboration>
Available 2008-09-16

Wall Communications Inc., 'A Study of Business Models sustaining the development of digital cultural content', CHIN, 2002
<<http://www.canadianheritage.gc.ca/progs/pcce-ccop/reana/model-study.pdf>>
Available 2008-09-16