Working with the Distributed National Electronic Resource (DNER):

Standards and Guidelines to Build a National Resource

Pdf : recommended for printing this document

Edited by: Catherine Grout, Caroline Ingram
Authors: Neil Beagrie, Hugh Buchanan, Tom Franklin, Catherine Grout, Brian Kelly, Caroline Ingram, Paul Miller, Greg Newton-Ingham, Andy Powell

Status: Version 1.0, for consultation
Feedback is welcome to: catherine.grout@kcl.ac.uk, or caroline.ingram@kcl.ac.uk

Date: February 2001

This document provides guidelines and practices critical to developing, managing, and delivering digital resources online. It will act as a source of information for projects and services involved in the development of the Distributed National Electronic Resource for learning, teaching and research. It will also provide a point of reference and information for the variety of stakeholders interested in the development of the DNER and for those interested in sharing or contributing content to it.

Adherence to standards plays an essential role in improving access to the information resources accessible online. Without the implementation of agreed approaches throughout the DNER, the aspiration of timely and usable networked information for use in education will not be realised.

Contents

1. Introduction to the DNER

1.1 DNER Goals

The DNER is the networked environment that provides users in further and higher education with a range of digital collections and advisory services to support their use. Its cornerstone is the development of a managed and mediated environment for accessing quality assured information resources on the Internet.

These resources include textbooks, journals, monographs, theses, abstracts, manuscripts, maps, music scores, still images, geospatial images and other kinds of data, as well as moving picture and sound collections.

The UK further and higher education community stands to make significant gains from involvement in the design, provision, and dissemination of world-class resources, and representation of this community's needs and expertise on the international stage is an important aspect of ensuring effective interoperability, whether within the UK alone, or between the UK and more remote resources.

Across information providers as diverse as public libraries, museums, and data archives, access to related resources beyond the physical holdings of a single institution is becoming increasingly important. Under a number of different initiatives, from the National Grid for Learning (NGfL) to the DNER, attention is focusing upon enabling users, wherever they may be, to discover and access resources drawn from many different sources.

The DNER has the potential to provide real improvements to how we access educational resources and associated services. One of the principles behind the DNER is that it is not a centralised service, and does not rely on a single dedicated entry point. In fact it is possible, and even desirable, that end users in the education sector will already be using parts of the DNER without being directly aware of it. The overall aim is an environment integrated with and embedded in the fabric of learning, teaching and research.

For general information about the DNER

http://www.jisc.ac.uk/pub99/dner_desc.html

http://www.jisc.ac.uk/pub99/dner_vision.html

The DNER provides a range of digital collections, for a full listing

http://www.jisc.ac.uk/dner/collections/licensing.html.

These collections are also listed by subject in the following Resource Guides

http://www.jisc.ac.uk/subject/

DNER Services are listed at

http://www.jisc.ac.uk/dner/services/

As a complex and rich environment the DNER will, by nature, undergo a process of continuous evolution and development. Some would argue that educational resources have always been distributed throughout the libraries, museums and archives of the UK, and thus a nascent DNER has always existed. The DNER programme and other activities led by the DNER programme office are intended to cement the technical infrastructure of the DNER and also re-articulate the concept of a shared national information resource in an evolving network space. There is therefore an inherent challenge in this development.

1.2 DNER Development & Enhancement Programme 2000-2003

The JISC 5/99 call for proposals to develop the DNER for learning and teaching resulted in the allocation

of more than £13 million to 40 projects across the UK. The majority of funding (about £9 million) has been allocated to projects for the enhancement of existing collections and the development of materials for learning and teaching purposes (Programme Area B). The remaining £4 million has been provided to projects working to develop the DNER service environment, including new 'fusion services' (subject portals) and infrastructure services such as article discovery and delivery services (Programme Area A). In addition, a formative evaluation of the programme has been commissioned. An additional £2 million has been set aside to address immediate gaps in the existing and planned provision.

1.3 Aim of Standards and Guidelines for the DNER

This document is aimed at providing a source of information and a set of guidelines regarding the standards required to create digital resources for the DNER. It is also intended to provide practical information concerning standards and information architecture issues for the projects involved in the DNER infrastructure programme.

Furthermore, the document defines a baseline set of technical policies and standards necessary to achieve interoperability and seamless information flow across the range of resources supported within the DNER. Additional standards may be necessary to support specific requirements.

Future Updates

It is also important to note, that this is intended to be a living document. Standards and guidelines will be modified and added to in future, as technology and the information environment offered by the DNER further develops.

Terminology Used

Requirements to comply with the standards and guidelines presented here are expressed as either;

MUST, SHOULD or MAY

MUST – means the recommendation must be followed or an alternative approach agreed with the JISC/DNER Office.

SHOULD – means the approach recommended is seen as highly desirable though following it is not mandatory

MAY – means that the approach mentioned is an acceptable way of proceeding and is broadly compatible with emerging DNER standards

1.4 Intended audiences

This document is intended for two main audiences:

Firstly as guidance for those involved in the DNER development programmes.

Secondly as a key point of articulation of the concepts driving the DNER for the variety of stakeholders and interested parties to whom this development is of relevance.

NB A more specific document exists for *publishers and other content suppliers* interested in contributing content to the DNER; **Joining the DNER** this is available at:

http://www.jisc.ac.uk:9191/dner/collections/dner_interoperability.htm

DNER Programme Projects

These guidelines have been written specifically for the organisations and individuals involved in contributing either content or infrastructure to the current DNER development programmes.

1. Certain standards will not apply to particular projects. However, it is important that project managers understand not only those standards that directly impinge on their research and development, but also those that contribute to the broader vision of the DNER.

2. Adherence to recognised standards will facilitate access to distributed information resources. This issue is important within the context of the DNER, but is also independent of it.

3. Most of the standards are ***recommendations*** rather than ***rigid requirements***, as few can be rigorously enforced.

4. However, DNER project managers are requested to ***comply*** with the standards laid out in this document as closely as possible to ensure that resources are not wasted, and that the outcomes of the projects are useful in the context of the DNER.

5. Projects ***unable to comply*** with recommendations that directly affect their project are requested to submit their reasons to the JISC/DNER Office.

Key DNER Stakeholders

It is intended that with maintenance and updating these standards and guidelines will become the required set of standards for DNER content creation and development, and a key method of communication with a broad range of stakeholders.

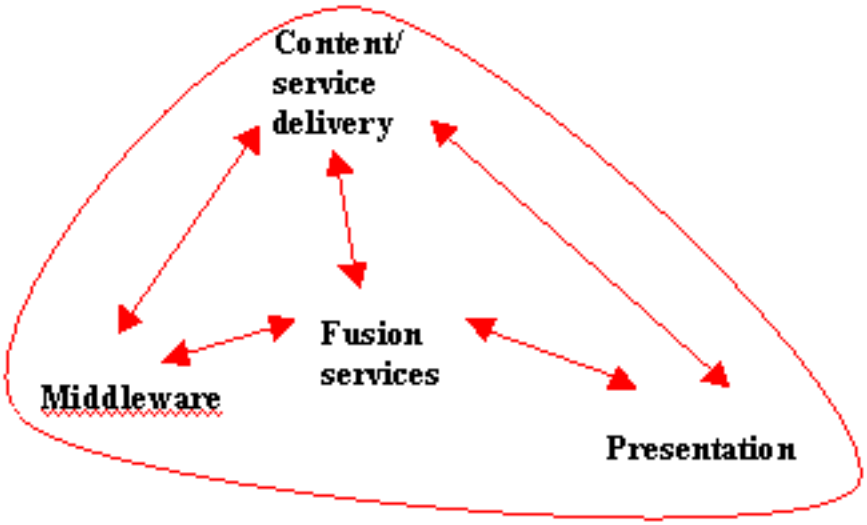
Important groups to whom this document should be of key interest and relevance are:

- Publishers and suppliers of digital content to the DNER
- International standards development bodies
- Library, museum and archival professionals and allied strategic and funding agencies
- Policy makers, information mediators, and creators and users of digital content in the further and higher education communities

1.5 DNER Architecture

The DNER will be developed over the next three years to become a more interactive, rich and increasingly seamless information environment.

The main features of the emerging DNER architectural vision are summarised in the following schema:



The DNER consists of :

Content delivery services who physically hold and manage digital collections for users

Middleware services that operate on the network and are designed to be used by the full range of DNER service providers. The ATHENS authentication service, (http://www.athens.ac.uk) is an example of Middleware

Fusion Services are machine mediated network services that fuse together content in useful ways. For example by discipline area, digital format or learning aim.

Presentation, is the layer where users gain access to the material. The user interfaces provided by the DNER presentation layer will be provided by a combination of content delivery services and some dedicated fusion services for example media or subject specific portals which may also be directly visible to end users.

The nature of the intended changes to develop the DNER environment over the next three to five years can be understood by considering the following diagrams:

Figure 1: Current DNER Architecture

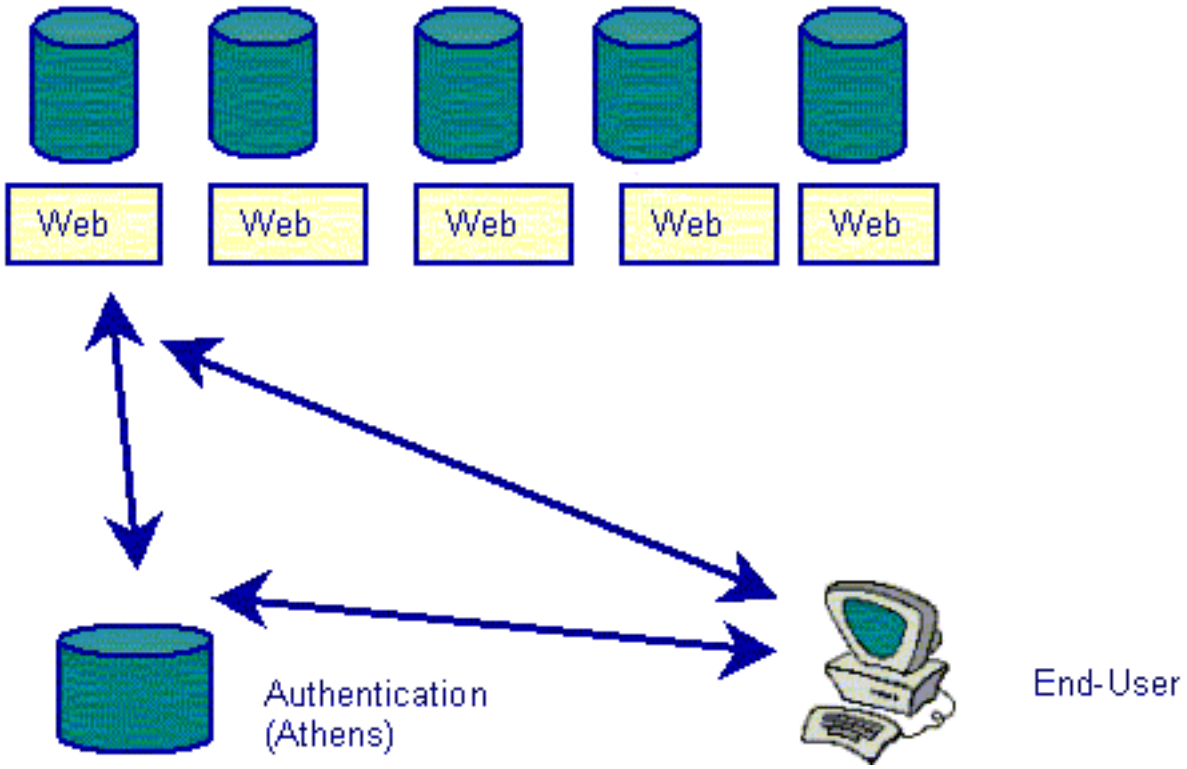


Figure 1: Current DNER Architecture

The diagram above shows the DNER as it is now, where a range of providers deliver content to end users via a number of separate web interfaces. There are some shared middleware services such as ATHENS authentication, but these are limited. Similarly subject gateways exist to undertake the job of providing subject specific resource descriptions, namely the RDN, (Resource Discovery Network http://www.rdn.ac.uk. However these gateways are providers of content, quality assured Internet resource descriptions, in their own right rather than providing the additional layer of fused content that is envisaged. Currently, there is limited interoperability between the digital collections being offered, this means that the goal of seamless access to the DNER has not yet come to fruition.

A more fully developed DNER architecture is expressed in **Figure 2.** below. This diagram is a current projection. Further research and development work is now underway to finalise the intended DNER architecture.

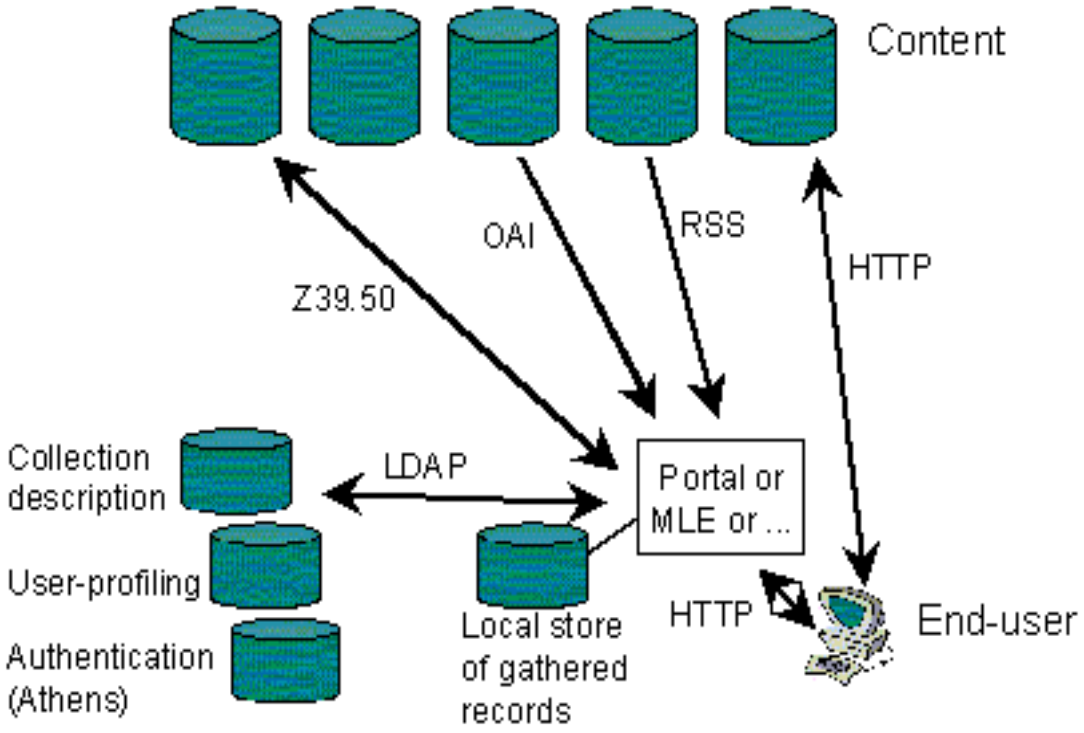


Figure 2: Future DNER Architecture

This document follows current thinking about the high-level DNER architecture, describing the standards that will enable content delivery services, middleware services and presentation services (fusion services/portals) to interoperate. The network services in each of these areas make up the DNER 'information environment'. Some services stand alone, however many will need to work together.

In order to do so they need to expose content and functionality in various ways as **Figure 2** indicates. Typical interfaces currently include HTTP, Z39.50, LDAP and the Open Archives Initiative (OAI) Harvesting Framework. These interfaces all allow the exchange of data, based on standard interchange formats such as MARC, RDF and XML. They enable the delivery of shared services.

The DNER has commissioned a study to investigate architectural models for such a network of services, validating recent thinking and augmenting it with more detail. In particular the study will look at the main functional areas identified in the high-level view and consider what the main components in each area might be and the interfaces through which they might communicate. The study will investigate standards for resource naming, persistent URLs, collection description, distributed searching, metadata sharing, distributed object technology, reference linking and user profiling, privacy and preferences.

It is important to appreciate in this context that the capacity to search across databases is growing. In the jargon, hubs are changing from being gateways of links and becoming portals. Increasingly they can be a part of managed learning environments (MLE's) leading directly to real data so that the user is able to reach easily into the resources that are needed without mediating a confusing and proliferating landscape of websites.

A key goal driving the development of the DNER architecture is the notion that end users require digital resources that are tailored to their needs and are presented as such. This need cannot be fulfilled without providing an information architecture that fuses content in a variety of useful ways, for example by

subject. e.g. Biology, or *medium*, e.g. moving images, or by *type of usage*, e.g. learning resources. Therefore the emerging architecture involves developing and user testing a new layer of "fusion services" which can either be mediated by machine or presented directly to the end user. This is an important feature of the information architecture which should lead to a more usable and relevant DNER.

1.6 DNER Rights Management

The JISC only funds the creation of resources that can be made freely available to higher and further education institutions in perpetuity. Therefore it is important to make this caveat clear when it becomes necessary to apply for the rights to use new materials. All resources **MUST** be available in perpetuity for learning, teaching and research purposes.

All resources **MUST** remain free at the point of use for the UK further and higher education communities.

DNER content, learning materials, or any other products developed as part of the Programme, **MUST not** infringe the rights of any third party in how they are created, managed, used, delivered or preserved.

Those creating content for the DNER are responsible for obtaining copyright clearance for content where required. Copyright clearance **MUST** include distribution of content via the DNER. Where material is in copyright it is important that the owners are keen to work with higher and further education in order to provide clearance in a timely manner.

Members of the DNER Development Programme will be provided with a set of detailed copyright guidelines which will include pro forma licences and copyright clearance letters which can be deployed wherever appropriate.

The requirements of data protection legislation **MUST** be complied with, so that personal details of individuals are not stored or made public without the knowledge and agreement of those concerned http://www.dataprotection.gov.uk/

Further Information

JISC Licensing Principles

http://www.jisc.ac.uk/pub00/licensing.html

Further JISC Copyright Links

http://www.jisc.ac.uk/pub/copyright/start.htm

BIDE, M., OPPENHEIM, C. AND RAMSDEN, A., 1997. *Copyright clearance and digitisation in UK Higher Education: Supporting Study for the JISC/PA Clearance Mechanisms Working Party*, Bath: UKOLN.

http://www.ukoln.ac.uk/services/elib/papers/pa/clearance/

Copyright Licensing Agency

http://www.cla.co.uk/

European Commission Legal Advisory Board, 1995.*The Information Society: Copyright and Multimedia*, [online].

http://www2.echo.lu/legal/en/950426/toc.html

JISC/TLTP Copyright Guidelines. London: LITC

http://www.ukoln.ac.uk/services/elib/papers/other/jisc-tltp/jisc.pdf

2. Content Creation

2.1 Background to this section

It has been proved that the application of standards and guidelines to digital resource creation projects will significantly enhance the widespread use and long-term viability of the content, so maximising the investments made and the potential that digital technologies can offer for teaching and learning.

This section establishes baseline standards for content creation and also provides a source of information regarding the process of content creation and how this should be managed. It does not aim to supplant the existing set of detailed and high quality publications in this area which deal with this subject in a more comprehensive manner. Links to further sources of in depth information are given where possible.

2.2 Content Selection

The process of deciding what to digitise anticipates all the stages of project management. Use of digital resources not only depends on the nature and importance of the original source materials, but also on the nature and quality of the digitisation process itself. Firstly on how well relevant information is captured from the original, and then on how the digital resources are organised, indexed, delivered to users, and maintained over time.

Selection decisions should be contingent with the aim of the DNER to develop a set of complementary and integrated resources for a particular discipline or across disciplines which include the full range of appropriate types of data and information.

The DNER has a **Collections Strategy** which provides a rationale for the resources that are being developed for inclusion in the DNER. http://www.jisc.ac.uk/dner/collections.html.

This will soon be accompanied by a Collections Development Policy which contains more detail about policies and procedures in this area.

Policies are also being developed to underpin the creation, management, use and preservation of resources in the strategic collecting areas with which the DNER is concerned, these are:

- E-books
- Journals
- Geospatial Data
- Primary Data
- Learning Materials
- Discovery Tools

- Images
- Moving Pictures and Sound

It is clear that the information resources considered suitable for acquisition or digitisation include a matrix of data types, kinds of material, and discipline areas. Material selected for creation may originate in digital and/or analogue form (e.g. microfilm and print). The resources with which the DNER is concerned are not restricted to such traditional resources as journals, monographs and textbooks, but extend to manuscripts, maps, music scores, and computer based learning packages. There is also an increasing need for the digitisation of sound recordings, diagrams, photographs, slides, and other materials for use in subjects ranging from medicine to catering.

The following criteria should be taken into account in the selection process:

- **Relevance to the actual or potential user needs** - materials should have sufficient intrinsic value to ensure interest and should be proven to have the capacity to significantly improve existing collections e.g. for enhancing the DNER for learning and teaching rather than research. An information resource should be selected for inclusion because it is, or will be, of value to teaching and learning in the UK higher or further education community.
- **Rights and permissions for electronic distribution should be securable within the parameters of the project and funding available** - where material is in copyright, preference should be given to working with organisations keen to co-operate with higher/further education, whether they are UK or international data suppliers, other key organisations funded within the HE/FE sectors, or non-commercial suppliers such as museums and galleries.
- **Quality of the original source material** – materials to be digitised should be of the highest quality possible, capture from original materials is generally preferable and especially where maximum information capture/fidelity to original sources is of importance.
- **Currency and timeliness** - many disciplines, particularly the sciences, require the most up-to-date information.
- **Depth of the existing collection in the subject area** - the new resource should be a worthwhile addition to the existing collection and the production of duplicated material should be avoided.
- **Scope and content** – all DNER content for the learning and teaching programme should show clear evidence and contextual training materials or metadata to demonstrate and exemplify its relevance to educational needs.
- **Metadata** – Resources created must be supported by the appropriate metadata; such metadata should describe the material comprehensively and coherently as well as providing clear pathways to the user attempting to navigate within the particular information environment.

For more information on standards for metadata see Section 2.7

Further Information

Guidance for selecting materials for digitisation

http://www.rlg.org/preserv/joint/selection.html

2.3 Preparation

Preparation is one of the most neglected areas of the digitisation chain, but if done well can help avoid many of the most common errors or quality defects that occur in large scale digitisation projects.

Preparation falls into two categories: preparation of originals, and preparation of the capture process. Where digitisation is *outsourced*, this work may be divided into, that done by the owner of originals before the materials are moved to the digitisation site, and that done by the operator of digitisation services before any capture takes place. If digitisation is taking place *in house*, both processes may be carried out by the same group. However it is important to build in the time to both prepare originals as well as set up the capture process before commencing.

Preparation should include bench marking. Benchmarking is the process of trialing digitisation techniques and methods and processes, in order to establish consistent and appropriate capture standards, and to establish workflow for the digitisation chain.

2.3.1 Benchmarking

Benchmarking is the process that **SHOULD** be undertaken at the beginning of any digitisation project. Benchmarking aims to set the standards to be used in the capture process. It will ensure that the most significant and useful information is captured. It requires:

- full knowledge of the main attributes of the original information
- understanding of present and recognition of future user needs
- exact guidelines on standards of acceptability that will meet all present and future requirements of the digital resource

Project managers **SHOULD** aim to produce a 'control' or trial set of digitised resources, either in-house, or as a requirement of the tender process in order to assess the efficacy of capture processes chosen. By carrying out benchmarking and following the three requirements given above the project manager should be able to:

- make a final decision about the standard of digitisation (in terms of resolution, level of detail, sampling rates, Esc)
- reconfirm or bring into question any decisions made at the content selection and preparation stages (especially whether the digitisation process can take place in-house, or needs to be outsourced)
- establish the hardware and software requirements of a project
- have a set of strict requirements arising from the results of the control exercise which can be written into any tender document and which can be used to make the decision regarding which out-sourcing organisation should be used
- establish workflow procedures and a delivery timetable for project planning.

Further Information

TASI Framework; image handling and preparation

http://www.tasi.ac.uk/framework/capture/imagehand.html

Preparation for capture of digital information

http://heds.herts.ac.uk/resources/papers/jidi_fs.pdf

Digitisation; Strategic and Management Issues

http://heds.herts.ac.uk/resources/Papers/HEDSITForum.pdf

Digitisation how much does it really cost?

http://heds.herts.ac,uk

2.3.2 Production Equipment and Facilities

It is difficult to provide proscriptive or detailed guidelines regarding selection of appropriate equipment for capture. The basic rule of thumb is to bear in mind fitness for purpose, longevity, and cost. It is also important to consider whether the resource should be digitised in house or whether this work should be out-sourced. Several factors govern this decision making process including, cost, viability of movement of originals, skills set available in house etc. The following sources of information should assist decision making;

HEDS, The Digitisation Process; Set up an in-house scanning unit or use a bureau?

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

JIDI (JISC Image Digitisation Initiative) Digitisation Feasibility Study

http://heds.herts.ac.uk/resources/papers/jidi_fs.pdf

VADS/TASI Guide to Good Practice, Section 3.

3.1 Introducing digital images and image creation
3.2 Digital images
3.3 Selecting image file formats
3.4 Image manipulation and imaging software
3.5 Image capture equipment: scanners
3.6 Image capture equipment: digital cameras

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

TASI Framework

http://www.tasi.ac.uk/framework/capture/hwandsw.html

RLG/DLF Guidelines; Guides to Quality in Digital Resource Imaging; Selecting a Scanner

http://www.rlg.org/visguides/visguide2.html

Overview of Scanners

http://www.gii.getty.edu/intro_imaging/11-Scan.html

How digital cameras work

http://www.cmospro.com/howdcw.htm

2.4 Workflow (or procedures management)

Workflow for digital content capture is the process of establishing and developing the chain of events which lead from an analogue work to a digital one. This may involve locating, un-shelving and preparing originals for digitisation, capturing work, creating metadata, post processing and storage. This work will probably be done by a team of people rather than one individual, often third parties, for example with digital imaging, a photographer or digitisation bureau may be an intrinsic part of this chain. The relationship between these parties, and how information, as well as digital or original objects are transferred defines the pattern of workflow. Smooth workflow where effort is not duplicated and which proceeds logically from one stage to the next is crucial.

It is likely that you will need procedures in place for managing the digitisation chain effectively and to keep track of the originals and digital surrogates as they are created. This process of tracking workflow can be time consuming. However it is an essential part of the process and should not be overlooked to the detriment of the end product but also potentially to the safety and condition of the original objects.

Further Information

JIDI/TASI Workflow Guidelines

http://www.ilrt.bris.ac.uk/jidi/workflow.html

VADS/TASI Guide to Good Practice, some dedicated sections on workflow

http://vads.ahds.ac.uk/guides/creating_guide/sect52.html

http://vads.ahds.ac.uk/guides/creating_guide/appendix2.html

2.5 Capture Standards

2.5.1 Introduction

The following section establishes some baseline capture standards for projects developing a range of digital content to be included in the DNER. It also has broader applicability for those creating and developing digital resources within the wider further and higher education community and beyond.

The following points should be adhered to during content capture:

- Standard and non-proprietary formats **MUST** be used wherever possible. They are more likely to have a good shelf life making the content produced suitable for archiving and preservation. (It is recognised that some areas, particularly time based media do not have stable nonproprietary formats currently available for capture). For further information see Section 2.5.3)

- Capture at the *highest possible resolution* **MUST** be achieved whilst bearing in mind the level of information content that it is appropriate to capture from the original. Some originals will contain a higher degree of useful content than others. It is also recognised that the digitisation process may be targeting particular information content and is therefore intended to emphasise this through the capture process.

Intended purpose is also a crucial driver. For example for archiving purposes, high resolution images may need to serve as reasonable substitutes in the event that the original item is lost or deteriorates. However for creation of learning materials to be delivered via networks in a variety of technological conditions, issues of practicality will undoubtedly surmount those of maximum information capture.

- Appropriate *web level surrogates* for use for on-line delivery **MUST** be generated where this is relevant to a project's dissemination goals. For example, high-quality digital images appropriate for preservation are large and unwieldy to use over a computer network, so projects need to consider the development of specifications for access-quality images. Lower resolution images - whose digital files should be smaller (in bytes) - can be more easily handled by personal or institutional computer systems. Such images would be lower in either spatial resolution ("dots per inch") or tonal resolution ("bits per pixel") or both, and **SHOULD** be derived, if possible, from the highest resolution images available.
- Overall the capture format chosen **MUST** be appropriate for the content and the mode of content delivery.

2.5.2 Still Images

Formats for Creation

Images **MUST** be created and stored using the TIFF or PNG format. These formats are non-compressed, and non-lossy and are therefore suitable for initial capture purposes.

Formats for Networked Delivery

Images **MUST** be provided on the Web as GIF (for line-drawings) or JPEG (for photographs) formats. Though PNG **MAY** be used as an alternative format to GIF.

For compressed images LZW compression **SHOULD not** be used. It is a copyrighted technology and is not suitable for long term storage and retention.

Capture Guidelines

Digital imaging techniques offer the opportunity to easily and quickly capture images or other archival materials. These can act as a means of improving access to materials or as a step toward preservation or both. In either case, projects developing digital content **SHOULD** refer to a set of practical guidelines that assist them in making decisions about image quality, presentation/display options, and storage formats.

Further Information

Digitisation Feasibility Studies, HEDS

http://heds.herts.ac.uk/resources/papers/jidi_fs.pdf

http://heds.herts.ac.uk/resources/papersF.html

Building Image Archives, TASI Framework

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

Guidelines for Image Capture, RLG/NPO

http://www.rlg.ac.uk/preserv/joint

Digital Imaging for Photographic Collections: foundations for technical standards

http://www.rlg.org/preserv/diginews/diginews3.html#com

Digital Imaging for libraries and archives,
Kenny A and Chapman S. (1996b), Cornell University Library (print based publication)

Creating Digital Resources for the Visual Arts; VADS, Section 3.

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

Guides to Quality in Visual Resource Imaging, DLF, RLG

http://www.rlg.org/visguides/

2.5.3 Time-based Media

Moving images, sound and music **SHOULD** initially be captured and stored using appropriate open or proprietary standard formats as outlined below: All projects **SHOULD** be aware that there maybe a medium to long-term need to recompress/re-encode media. With that in mind, projects should maintain a copy of all resources in a high quality format to enable re-purposing and recovery.

Moving Images

Encoding (digitisation): Moving Picture content **SHOULD** be encoded using either Apple Quicktime, Microsoft Windows Media Player or MPEG1, MPEG2, or MPEG4. Where codecs are not implicit, encoding should be done using codecs supplied free with the appropriate players and which enable use on all relevant platforms.

Formats for re-encoding: DV, miniDV, DVCam, DV Pro, DV Pro 50, digiBeta and BetaSP are all acceptable.

Sound

Encoding: Sound content **SHOULD** be encoded using either Apple Quicktime, Microsoft Windows Media Player or MPEG Audio Layer 3. Where codecs are not implicit, encoding **SHOULD** be done using codecs supplied free with the appropriate players and which enable use on all relevant platforms.

Sound formats for re-encoding: CD-Audio (44Khz @ 16Bits) or DAT (44Khz @ 16Bits or above) are recommended.

Practical guidance can be found at:

The JISC advisory service for Moving Images and Sound;

http://www.bufvc.ac.uk/

Guidance on encoding can be found at

http://www.codeccentral.com/

PADS; Guide to Good Practice - Creating Digital Performance Resources

http://www.pads.ahds.ac.uk/padsGGPPerformance

2.5.4 Text (e-books, journals, other texts)

Formats

Text based content **MUST** be created and managed in a structured format suitable for delivery as HTML. In most cases storing text-based content as HTML, XHTML or XML, either in plain files or within a database of some kind, will be the most appropriate option. Where HTML is chosen as the storage format, the latest version recommended by the World Wide Web Consortium (W3C), currently HTML 4, or the latest version of XHTML, currently 1.0, **MUST** be used. HTML and XML documents should be validated against a published Document Type Definition (DTD).

Networked Delivery

Text-based content **MUST** be delivered as HTML or XHTML in most cases, though the use of other XML DTDs may sometimes be appropriate. Where HTML is chosen as the delivery format, the latest version recommended by W3C (currently HTML 4) or the latest version of XHTML (currently 1.0) **MUST** be used. HTML and XML documents **MUST** validated against a published DTD. In some cases, delivery in proprietary formats such as PDF, RTF or Microsoft Word may be appropriate.

Capture Guidelines

Oxford Text Archive; Creating and Documenting Electronic Texts

http://ota.ahds.ac.uk/documents/creating/

2.5.5 Discovery Tools

Please refer to **Section 2.6** Cataloguing and Metadata Guidelines

2.5.6 Geospatial Data

Special factors

Geospatial data can be held as raster, vector or object-based data.

There are factors that apply to geospatial data that do not apply to other forms of data. These factors **MUST** be documented in any metadata, and where appropriate included with the data itself. Among these factors are:

- scale and resolution of source information
- projection and spheroid (e.g. GPS data uses WGS84, Ordnance Survey-based data uses the Airy Spheroid and a Transverse Mercator projection)
- co-ordinate reference system (e.g. Ordnance Survey based data uses National Grid, satellite

imagery uses a locally defined reference system)

- data about spatial relationships (topology)
- storage format and version number

Raster

Raster storage has much in common with still image capture (see 2.5.2). In addition, if you convert aerial photography or satellite imagery you **MUST** also record the meaning of each of the pixel values (e.g. colours) that occur in the image.

Vector

When vector data is captured, each line, point or area feature has a Feature Code attached to it. The meaning of these codes **MUST** be clearly documented. Similarly when features have textual or numeric attributes, then the meaning of any codes used for those attributes must also be recorded.

For storage formats, national standards such as BS7567 (NTF) are publicly documented, but their support by software packages is patchy. Industry standards such as Arc/Info Export files, ArcView Shape files and MapInfo files are proprietary, not publicly documented, and not always easy to transfer between different systems. Therefore once data has been captured, continued attention **MUST** be given to ensure that it remains usable.

Object

The use of object data is still quite novel, and therefore may not yet be suitable for data capture purposes. No stable standard formats have yet emerged. One of the most promising is GML, which is based on Extensible Markup Language (XML).

Further Information

The ADS Guides to Good Practice provide detailed guidance on some of these aspects:

- *GIS Guide to Good Practice*
  http://ads.ahds.ac.uk/project/goodguides/GIs/
- *Archiving Aerial Photography and Remote Sensing Data: A Guide to Good Practice*
  http://ads.ahds.ac.uk/project/goodguides/apandrs/
- *The NCGIA Core Curriculum project aims to create Web resources for a wide range of topics, including geodata capture:*
  http://www.ncgia.ucsb.edu/giscc/

Technical Committee 211 of the International Standards Organisation (ISO) is currently drafting a wide range of standards on geospatial data which may prove useful.

2.5.7 Learning Materials

IMS Specifications

Learning technology has only recently begun to be systematically standardised and there have been a number of initiatives to achieve this. Among these are the AICC (Airline Industry CBT Committee) which has the longest history as aeroplane manufacturers understood that the aeroplanes last 20 years

while computers rarely more than 5. Since then there has been the EU funded ARIADNE initiative (no relation to the eLib project) and now there is IMS (global see http://www.imsproject.org/), ADL (US see http://www.adlnet.org/), PROMETEUS (EU see http://www.prometeus.org/) and IEEE LTS (global see http://grouper.ieee.org/p1484/).

All the specifications that these bodies are producing are converging and it seems most likely that standards will emerge from the IMS work.

IMS has broken the work down into a number of areas and has developed (or is developing) specifications in each of these. The intention is that they should then go via IEEE to become standards in much the same way that W3C specifications only later become standards The areas are:

Projects **MUST** be aware of the following and implement where viable;

- The IMS Metadata specification to describe learning materials. NB this is built on the Dublin Core.
- The IMS Question and Test Interoperability Specification to write and describe assessment materials (questions or tests)
- The IMS Content Packaging Specification which covers the description, structure, and location of online learning materials and the definition of some particular content types.
- The IMS Learner Information Packaging Specification is a collection of information about a Learner (individual or group learners) or a Producer of learning content (creators, providers or vendors). The IMS Learner Information Packaging (IMS LIP) specification addresses the interoperability of internet-based Learner Information systems with other systems that support the Internet learning environment.

Projects **SHOULD**

- Use the other IMS Specifications as they are approved.

2.6 Cataloguing and Metadata Standards

2.6.1 Introduction

In managing digital collections and supporting their use, the need for accurate information or metadata (structured data about data) is essential if we are to understand the identity, context and usefulness of a collection and locate it in a networked environment.

The need to improve the effectiveness of searching for information resources on the Internet has prompted the development of simplified metadata standards that can be used at relatively low cost. (The term "metadata" refers to data which relates to one or more information resources, supporting their discovery or management).

These metadata standards are intended to improve the general efficiency of searching for information on the Internet. However it is important to note that they do not supplant the need for good and consistent cataloguing of digital collections. The creation of standards compliant metadata is crucial in order to provide usable and sustainable resources.

This section gives an over view of the metadata and cataloguing standards relevant to the digital collections within the DNER, its also summarises some of the main issues concerning good practice in

the generation, structuring and development of useful metadata.

2.6.2 Good Practice for Cataloguing and Metadata Creation

For the purposes of this document, cataloguing and metadata creation refers to the process of creating structured descriptions that provide information about any aspect of a digital resource.

- This information may include technical information about both the *digital entity*, e.g. a scanned image, but also the photograph or *original object* from which the digital entity is derived.
- This information is also likely to describe the *whole of a resource* e.g. the whole digital or physical collection as well as its component parts.
- Other metadata is likely to be that which describes *the process of digitisation*, which documents the decisions that have been made during the resource creation process, and provides a comprehensive and authoritative way of referring to the finished product and its contents. (Types of specific *process* metadata may be *administrative* metadata, *technical* metadata and *preservation* metadata)

There are some important general points to bear in mind when creating metadata and cataloguing digital resources:

Projects **MUST**

- **Wherever possible and practical adopt an existing standard** for digital resource description and use recognised standards for terminology and relevant controlled vocabularies, see 2.7.4, and 2.7.5;
- **Spend time and attention over the design of the database or other structure which holds the cataloguing information and metadata.** Straightforward database structures are generally preferable (see 2.7.2);
- **Make sure this data structure is thoroughly documented** in order that you and other third parties can understand your rationale. For example, if you are using a database it is important to document how the fields you have selected are intended to function, and how the tables relate to each other;
- **Pay equal attention to creating all the types of metadata described above,** e.g. *technical metadata* is needed as much as metadata describing the subject or content of a resource. With the passage of time without the metadata which tells us about the digital object we will may not be able to identify the format of a resource. This also relates to the concept of *preservation metadata*, e.g. metadata which provides the information about how to maintain a resource in the long term;
- **Wherever possible provide detailed guidelines for cataloguing and metadata creation**. These should explain any rules of how fields should be populated or how different categories of information should be created;
- **Metadata should always be internally consistent.** E.g. consistent spelling, and use of classification schemes or terminology.
- **To make sure quality is maintained cataloguers should be trained** before being permitted to create metadata.

Projects **MUST NOT**

- **Create their own standards for metadata**, or substantially amend or add to an existing standard

for internal use without providing an adequate rationale for this process or documenting what changes have been made.

- **Allow untrained staff to catalogue or create metadata**
- **Keep multiple and varying copies of the same database or metadata**
- **Fail to build in automatic or other checks to ensure the consistency and quality of metadata**

2.6.3 Database Structures: selection and Implications

Designing a suitable data structure is the key to ensuring that the creation of metadata operates smoothly. Creating a data structure is the stage where relationships between different aspects of metadata, e.g. object and digital surrogate, technical metadata and content based metadata are defined, as well as the relationship between the intellectual concepts described within the metadata. In a database these relationships are often separated into tables within which information about a single entity or concept is focused. The following considerations should be borne in mind when creating a data structure.

Projects **SHOULD**

1. **Reduce complexity if possible.** In general, complex and multi-layered data structures should be avoided. For example a database which has several tables, with different hierarchies of entities within it and different relationships between entities (for example one to many) is inherently more difficult to manage. This is particularly true given that the need will arise to migrate the database through changes in software, platform or staff.
2. **Do not create extraneous metadata that you do not need.** Consider carefully the intended purpose for the information you are including.
3. **Use a database that supports open standards.** Select a database which allows you to import and export the metadata into a range of commonly used formats that are independent of a particular platform rather than proprietary. ASCII text based formats are the safest to use for the purpose of exporting and importing data, and your database software should support their use. For example CSV, Comma Separated Variable Format, is a commonly used text format for export of data from a database. ASCII based formats are also more *preservation* friendly. For example, MS Access and Filemaker Pro both support the import and export of ASCII based formats.
4. **Document your database design and structure.** As above make sure you store documentation about your database design and its rationale, this will allow others to make sense of it.

Bear in mind that for relevant DNER Programme Projects, third parties e.g. JISC Data Centres and Services will need to be able to make sense of your data in order to integrate it within the DNER. Following the practices described above will make this process work more efficiently.

Further Information

Metadata for the Masses, Paul Miller
(A useful primer to metadata and Dublin Core. Please note that technical information is now out of date)

http://www.ariadne.ac.uk/issue5/metadata-masses/

Forum for Metadata Schema Implementers

http://www.schemas-forum.org/

VADS Guide to Good Practice, Standards for Data Documentation,

http://vads.ahds.ac.uk/guides/creating_guide/sect41.html

2.6.4 Digital Resource Description Standards

It is worth noting that it is impossible to recommend a single resource description standard in this context that is sufficiently generic to provide commonality across the diversity range of digital resources under question, but at the same time capable of being sufficiently specific to cater for all the needs you may have of it.

In general terms the more generic a standard is, the more work you may need to do to map it to your needs.

It is also important to note that the information architecture of the DNER is still evolving. In the next months it is expected that more detailed definitions of metadata standards will emerge to support interoperability in the longer term. However at this stage it is important to consider use of the following standards.

General Standards:

Dublin Core

Prominent amongst general standards for the creation of metadata is the *Dublin Core* metadata set (http://purl.oclc.org/dc/). Dublin Core is being developed as a generic metadata standard for use by libraries, museums, archives, government and other publishers of online information. This standard may be applied broadly to citation and full text descriptions, and may support interoperability between a range of schemes.

Projects **MUST**

EITHER use the Dublin Core as a basis for their metadata standard **OR** provide a mapping to Dublin Core metadata which explains the relationship between the metadata standards they are using and the Dublin Core. This mapping should preferably relate to both *collection level* and *item level* metadata.

Implementations of Dublin Core

Although the core standard is now established, it is recognised that the Dublin Core is still under some development, in relation both to its semantic aspects (rules for the content of the fields) and the syntax (rules for structuring and expressing the fields).

However the following guidelines should be useful in indicating a standard way to implement Dublin Core:

http://purl.oclc.org/dc/documents/wd/usageguide-20000716.htm; *Using Dublin Core*

Collection Description Standards

One of the keys to the interoperability of digital resources within the DNER is having standardised ways

to provide metadata about the digital collections which exist within it. As a result we are recommending that participants employ the results of the following work when considering how to describe digital collections.

Projects **SHOULD** either use or map to the *RSLP Collection Description* schema.

Information can be found at:

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

Full Collection Description Schema at:

http://www.ukoln.ac.uk/metadata/rslp/schema/

Specific Standards

While it is important to use or map to metadata standards which have a broad applicability as above, it is probable that you will need to consider using standards which have specific relevance to the community you represent, e.g. archival, museum, educational etc., the subject areas relevant to you, e.g. medicine, art or archaeology and furthermore, there may also be metadata standards geared up to the kind of resource you are creating, e.g. visual as opposed to textual resources. Below are some examples of these categories and the standards applicable to them.

Projects **SHOULD** take account of relevant standards specific to their areas when creating metadata.

The following gives an indication of the range of standards currently in use:

- **Domain Specific Standards (Subject Areas or Practice Communities)**

Visual Arts

For a general description of relevant standards see the Visual Arts Data Service at:

Section 4: Standards for Data Documentation in the Visual Arts

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

Museums and Cultural Heritage

ICOM-CIDOC and Getty Information Institute

Developments in museum and cultural heritage information standards

http://www.cidoc.icom.org/stand1.htm

Museums and cultural heritage information standards resource guide

http://www.cidoc.icom.org/stand2.htm

Visit the standards page at the MDA, for links to the Spectrum Data Standard and other relevant work.
http://www.mda.org.uk/index_s.htm

Archives

ISAD(G): General International Standard Archival Description

http://www.ica.org/cgi-bin/ica.pl?04_e

- **Format/Process Specific Standards**

Web Pages

Resource Discovery Network Cataloguing Guidelines

http://www.rdn.ac.uk/publications/cat-guide/

Images and Visual Resources

Visual Resources Association Core Record

http://www.gsd.harvard.edu/~staffaw3/vra/vracore3.htm

JIDI (JISC Image Digitisation Initiative) Guidelines for Digital Image Metadata

http://www.ilrt.bris.ac.uk/jidi/metadata.html

Technical Metadata for Still Images; NISO Draft Standard

http://www.niso.org/commitau.html

Text

Oxford Text Archive; Creating and Documenting Electronic Texts

http://ota.ahds.ac.uk/documents/creating/

Guidelines for electronic text encoding and interchange

http://www.hcu.ox.ac.uk/TEI/Guidelines/

Preservation Metadata

CEDARS Project Preservation Metadata work

http://www.ukoln.ac.uk/metadata/cedars/AIW01.html

The RLG Working Group on Preservation

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

2.6.5 Terminology Standards and Controlled Vocabularies

Terminology standards and controlled vocabularies are a way of controlling the terms we apply during the process of metadata creation. Their application allows us to ensure that terminology is both internally consistent, e.g. consistent within a given set of metadata but also that it is consistent with the terminology and vocabulary used by others who have also adopted common standards in this manner.

Where possible, and where these are available:

Projects **SHOULD** use a relevant terminology standard or controlled vocabulary.

Examples are:

TGN: Thesaurus of Geographic Names

http://shiva.pub.getty.edu/tgn_browser/

AAT: Art and Architecture Thesaurus

http://shiva.pub.getty.edu/aat_browser/

Union list of Artists Names:

http://shiva.pub.getty.edu/ulan_browser/

wordHOARD

An index to terminology projects and resources relevant to the Museums Sector

http://www.mdocassn.demon.co.uk/wrdhrd1.htm

DDC21: Dewey Decimal Classification Schema

http://www.oclc.org/dewey/

MESH: Medical Subject Headings from the National Library of Medicine

http://www.nlm.nih.gov/mesh/

2.6.6 Subject/Resource Classifications for DNER Navigation

It is important for participants in the DNER to bear in mind that on-going work is taking place to look at a high level set of subject classifiers, e.g. based on higher and further education classifications of subjects taught in universities and colleges as well as a set of classifications for describing the digital resources the DNER provides access to.

The DNER team are currently commissioning a DNER collections subject mapping exercise, to look at mappings to both subject headings and resource types for collections management purposes. The results of this exercise should help to inform the process of providing classifications of resources to assist navigation.

Furthermore the JISC is jointly funding the HILT Project in which the DNER team is actively participating.

http://hilt.cdlr.strath.ac.uk/

HILT - High-Level Thesaurus is a one year project jointly funded by the RSLP and JISC. The purpose of the project is to study and report on the problem of cross-searching and browsing by subject across a

range of communities, services, and service or resource types.

2.7 Submission of content for the DNER

2.7.1 Introduction

This section is concerned with how projects in the DNER development programme should submit their digital collections or other deliverables for inclusion in the DNER.

Other organisations or individuals interested in providing content that adheres to DNER standards or with the interoperability of their content with that of the DNER should refer to this document more broadly to understand the baselines standards that are required.

The JISC currently funds a number of Data Centres and services who manage and archive collections for use by FE and HE. These services make up the core DNER information environment and take responsibility for the delivery, management and preservation of resources for learning, teaching and research. Relevant Services are listed at:

http://www.jisc.ac.uk/dner/services/

2.7.2 Working with content delivery partners

All projects creating digital content will be allocated a delivery partner, projects will be informed shortly concerning who they are working with. Projects are not precluded from delivering content themselves via the Internet or otherwise and are encouraged to do so. However the priority of the DNER programme is to deliver resources within the DNER framework and information environment, therefore project funds and resources **MUST** be directed toward working with delivery partners as a primary aim.

2.7.3 Quality Assurance

It is imperative that all projects undertake QA checks at various stages, regardless of whether the material has been produced in-house or via an outside source.

Projects **MUST** provide internal quality assurance checks, these checks are:

- Checking that DNER standards and guidelines have been adhered to. If standards are not adhered to in a particular scenario, a rationale must be provided to the JISC/DNER Office. JISC/DNER Office is likely to be supportive of reasonable requests.
- Ensuring that results are internally consistent, up to standard and fit for intended purpose. This can be achieved by spot checking of metadata or other digital content or through building in automated methods to check consistency, e.g. picklists to control terms to be added from thesauri etc.
- Providing documentation to describe the quality assurance methods that have been deployed

Projects should be aware that further assistance and some external quality assurance testing will be organised by the JISC/DNER Office.

Further Information

JIDI/TASI QA Guidelines

http://www.tasi.ac.uk/framework/collections/qassurance.html

http://www.tasi.ac.uk/framework/collecions/jidiqa.html

RLG/DLF Guides to Quality in Visual Resource Imaging 4. Measuring Quality Of Digital Masters

http://www.rlg.org/visguides/visguide4.html

RLG/DLF Guides to Quality in Visual Resource Imaging 3. Imaging Systems: the Range of Factors Affecting Image Quality

http://www.rlg.org/visguides/visguide3.html

2.7.4 Delivery Medium and Method

For members of the DNER development programme a suitable delivery medium e.g. Dat tape, CD ROM, FTP etc. will be defined in consultation with allocated delivery partners.

Other than in exceptional circumstances the JISC/DNER Office will NOT undertake receipt of submitted materials. These will be delivered straight to the delivery partner concerned.

Submission Documentation

Resources created as part of the development programme are useless without supporting documentation which describes how and why they have been created and what they contain.

The Arts and Humanities Data Service and the the Data Archive have provided guidelines for submitting digital content for archiving and preservation. These provide a useful starting point regarding the kind of documentation that should accompany any resource that is submitted.

Managing Digital Collections: AHDS Policies, Standards and Practices; Section 2.9 Guidelines for Depositors

http://ahds.ac.uk/public/srg.html#guidelines

The Data Archive

http://www.data-archive.ac.uk/depositingData/howtoDeposit.asp

3. Preservation & Records Management

3.1 Background to this section

This section first deals with the range of strategies and approaches which will help to ensure important digital resources do not become inaccessible prematurely. Many constitute a relatively modest investment compared to the initial costs of creating the resource, which are often substantial. They can also represent significant cost savings longer term. In any event, failure to commit resources to managing digital resources throughout their lifecycle will inevitably result in their loss and/or costly restoration so investment in strategies to prevent this is eminently justified.

Maintaining access to digital resources over the long-term involves inter-dependent strategies for preservation in the short to medium term based on securing, storage media, content and documentation, and computer software and hardware; and strategies for long-term preservation to address the issues of

software and hardware obsolescence. This section is therefore divided into two parts: the first dealing with storage and maintenance of digital resources, the second with strategies for their long-term preservation.

A preservation strategy for digital resources is most effective if it addresses the full life-cycle of the resource allowing the greatest efficiencies between data creation, preservation and use. This section should therefore be read in conjunction with related sections of the Guidelines.

There is no single definitive solution which can be applied for the preservation of any digital resource. However an approach which is based on good management practices introduced as early as possible in the lifecycle of a resource, will safeguard the initial investment and facilitate authorised access at least for the short to medium term.

Records management is an important aspect of managing digital resource creation and is essential for effective project management. Efficient organisation of project records and files will assist the efficiency of projects. A records management consultancy is currently in progress and will be providing records management guidelines for JISC projects. These will be added to these guidelines in April 2001.

3.2 Storage and Maintenance

3.2.1 Storage Media and File Formats

Media

Storage media are an important consideration and their selection will be based on a combination of factors including storage capacity, data transfer rates, the manufacturer’s market presence and reputation, and the IT infrastructure of the organisation. Different requirements and criteria may apply to storage media for access and archive copies. Depending on the needs of the organisation and the media, it may be necessary to create both preservation and access copies and to have strategies for each.

Magnetic media provide a versatile and cheap storage medium. Digital Linear Tape (DLT) is frequently recommended for archival storage because of its high storage capacity and good data transfer speeds. Magnetic media are constantly evolving and in addition to frequent changes in devices, manufacturers often undertake an almost constant evolution of production processes. It is important to be aware that faults in manufacture can occur and to make appropriate safeguards in backup and recovery procedures. Magnetic Media **SHOULD** be of high quality and purchased from reputable brands and suppliers.

Optical storage media such as CD-ROM, CD-R, and DVD use laser light to read from a data layer and are an increasingly popular method of storage. As with magnetic media, optical media have been subject to a constant process of evolution and changes in manufacture. Smaller storage capacity and slower data transfer rates either writing to or from the media can be issues to consider. The quality of the media, a reputable source, and appropriate handling and storage environment will all affect its longevity. The use of light sensitive dyes means some CD-R's are less stable than CD-ROMs and more concerns have been raised over their use as archival media. Optical Media **SHOULD** be of high quality and purchased from reputable brands and suppliers.

In addition to the media it is important that attention is paid to the recording and access devices such as tape drives. These **SHOULD** be of good quality and well-maintained. Problems with the access devices e.g. head/media crashes are one of the most common causes of damage to magnetic storage media.

Institutions **SHOULD** also:

- follow the media manufacturer’s recommendations for labelling. Do not place labels on optical disks and/or mark using a pen or pencil.
- minimise handling and use of archival media and/or record number of accesses/use and implement appropriate refreshing.

File Formats and Standards

As with storage media there is a diverse range of formats (e.g. Word, TIFF) in common use. The purpose of this section is not to provide a detailed or exhaustive list of current formats for different media types but to draw attention to the broader implications of file formats for their application, and implications for preservation. There are a number of excellent sources of more detailed advice on file formats and these are detailed further in this chapter.

File formats are subject to similar rapid obsolescence and evolution and the process of selection and assessment of options for preservation is largely one of risk reduction. Use of file formats which have been well documented, have undergone thorough testing and are nonproprietary and usable on different hardware and software platforms minimises the frequency of migration and reduces the risk and costs in their preservation. Similarly utilising formats which have been widely adopted minimises risk as it is more likely that migration paths will be provided by the manufacturers and a degree of 'backward compatibility' will be available between versions of the file format as it evolves. It is important to note that backward compatibility is rarely maintained for more than one or two previous versions and that the "window of opportunity" to migrate is therefore relatively brief.

Although such nonproprietary formats can be selected for many resource types this is not universally the case. For many new areas and applications, e.g. Geographical Information Systems or Virtual Reality only proprietary formats are available. In such cases a crucial factor will be the export formats supported to allow data to be moved out of (or into) these proprietary environments.

In considering storage and preservation it is helpful to recognise that it can be a desirable strategy to distinguish between formats (or versions) used for archiving and access on the basis of different requirements. E.g. it would be appropriate to store a high resolution image as a TIFF master file (archival format), but to distribute the image as a JPEG file (access format) of smaller size for transmission over a network. It would not be appropriate to store the JPEG image as both the access and archival format because of the irretrievable data loss this would incur.

The speed with which many file formats evolve and the degree to which even well documented standard formats can be extended by proprietary additions or modified/adapted for specific applications by users also has significant implications for preservation, and in particular for good preservation metadata and system documentation.

Institutions **SHOULD**

- Use 'open' nonproprietary, well-documented file formats wherever possible.
- Alternatively utilise file formats which are well-developed, have been widely adopted and are 'de facto' standards in the marketplace.
- Identify formats acceptable for the purposes of transfer, storage and distribution to users (these

may be distinct).

- Minimise the number of file formats to be managed as far as is feasible/desirable.

Do not use encryption or compression for archival files if possible.

Compression and Encryption

File compression algorithms can substantially reduce file sizes and have been widely used in document or image transmission. Compression can either be losses or lossy (with data loss but often higher levels of compression). Although appropriate in many cases for access and user copies, compression adds additional complexity to the preservation process and SHOULD not be used for storage of archival files. With current increases in storage capacity and reducing costs it is also less necessary. For some very large files e.g. digitised video, compressed formats may be the only viable option however for capture, storage and transmission.

In a similar way encryption is increasingly prevalent either to ensure that sensitive data is read only by the recipient or to ensure a digital product can only be used by an authorised user. Encryption also adds to the complexity of the preservation process and **SHOULD** be avoided if possible for archival copies. This may require strict implementation of physical and system security procedures for the archive of unencrypted files, or archival access to encryption keys.

3.2.2 Maintenance

Digital data is very secure providing appropriate management is followed. However if this management is not in place digital data is easily corrupted or destroyed. Institutions **MUST** implement a range of procedures both during and after projects to ensure the digital resources they are creating are not subject to loss or damage and will be suitable for future access and use. Digitisation projects may generate substantial storage requirements and the need for specialised facilities. These may be available in-house, as part of a consortium, or from third party services.

Some of the key procedures are:

Backups and Disaster Recovery Planning

Institutions **MUST** ensure that investment in digitisation is safeguarded through the implementation of appropriate regular backup and disaster recovery procedures. Multiple copies of the data stored both on and offsite reduce the risk of loss or damage and procedures for regular updating of backups **MUST** always be in place. Depending on local institutional arrangements, it may be possible to automate backups over a network. You **SHOULD** ideally write at least two backup copies, to two different media to guard against faults introduced by media's suppliers into their products.

Media refreshing and reformatting

Media refreshing and reformatting are essential management components for all digital media to avoid media degradation and to facilitate longer term preservation strategies. Archive copies will be periodically refreshed onto identical media to address media degradation and impermanence and are reformatted/transferred to new storage media as storage technologies change. Media refreshing and reformatting **MUST** take place:

- within the minimum time specified by the supplier for the media's viability under prevailing environmental conditions;
- when new storage devices are installed;
- when an audit discloses significant temporary or read "errors" in the resource.

Environmental conditions

Appropriate environmental conditions will increase the longevity of digital storage media and help prevent accidental damage to a resource or its documentation. Institutions **SHOULD**

- Follow relevant guidance on environmental conditions for storage media
- Prohibit smoking and eating in the storage and work areas.
- Store away from direct sunlight.
- Provide additional protection in the form of enclosures for media.
- Provide storage facilities which minimise the threat from natural disasters such as fire and flood or to magnetic storage media from magnetic fields.
- Ensure any non-digital accompanying materials (e.g. codebooks, operating instructions) are also stored in appropriate environmental conditions.
- Establish guidance and procedures for acclimatising magnetic tape if moving between significant variations in temperature (e.g. tapes moving from very cold external conditions should not be used before being acclimatised to warmer internal conditions).

Audit

Audit procedures provide reassurance that the resource has not been inadvertently or deliberately changed following refreshment and/or migration procedures and to check the readability and integrity of the data over time. Employ quality control procedure such as bit/byte or other checksum comparisons with originals to ensure the authenticity and integrity of items after media refreshing. Institutions **MUST**

- Check media periodically for their readability. Such checking may be conducted automatically in mass storage systems according to parameters set by system operators.
- Check the integrity of data files periodically using checksum procedures. Such procedures may be implemented automatically in mass storage systems according to parameters set by system operators.
- Employ appropriate security systems and procedures to protect the authenticity of items in your holdings (see security below).

Security

Data security procedures are essential for preservation and integrity of the resource by preventing alteration or loss. It is important to note that not all digital resources will require identical levels of security. Guidance on levels of security can be found in BS 7799 Information Security Management. To safeguard resources institutions **MUST**

- Establish a disaster recovery plan.
- Control access to storage facilities and processing areas.

- Use passwords and user ids, and other network security procedures.
- Define system and area access privileges for staff.
- Assign specific staff responsibilities for data security and storage facilities.

3.3 Preservation Strategies

Digital preservation differs from the preservation of traditional analogue materials. Archives traditionally preserve physical objects which carry information. Digital preservation is concerned with preserving information regardless of the object on which that information is stored. This is because software and hardware used to access the information change rapidly and become obsolete and the physical media on which digital data are stored are impermanent.

The DNER aims to develop and maintain resources over the long-term and this involves interdependent strategies for preservation in the short to medium term based on securing the computer systems, storage media, data and documentation; and strategies for longer-term preservation to address the issues of software and hardware obsolescence. Common strategies for long-term preservation can be summarised as follows:

- **migration** (data is stored in software independent formats and migrated through changing technological regimes)
- **technology preservation** (data is preserved along with the hardware and/or software on which it depends)
- **emulation** (the look, feel, and behaviour of a resource is emulated on successive hardware/software generations)

It is envisaged that most projects will only be responsible for the creation and maintenance of digital resources for the lifetime of the project. Projects **MUST** make arrangements for long-term preservation and access with an appropriate repository. The repository will then be responsible for implementing long-term preservation strategies and procedures. Further guidance on preservation strategies is available from JISC Digital Preservation Focus (Beagrie and Jones forthcoming).

3.4 Records Management

Good records management practice will support the efficient management of projects and ensure that legal and business requirements are met. JISC project staff **MUST** be aware of the records management requirements and practices of both JISC and their host institution.

The JISC funded *Study of the Records Life Cycle* by Elizabeth Parker published in 1999 provides guidance on retention and good management practice for local institutional records and has been adopted by many institutions. The study is available on the JISC website at http://www.jisc.ac.uk/pub01/records_lifecycle/.

In addition during 2000-2001 the JISC commissioned the study *Records Management in JISC funded Project-Based Programmes* by Kelvin Smith specifically on its own requirements for records. This report will shortly be available on the JISC website .

The following extract based on that report provides guidance to staff on filing and managing project records.

Summary

1 This strategy takes account of paper and electronic records. Paper records should be kept in conventional file covers, arranged in subject order in line with the file structure and classification guidelines described below. Papers should be secured in the file jackets in chronological order. This will enable easy location and retrieval of documents and promote preservation of the information. All reports generated by JISC programmes (with the exception of administrative and progress reports) should be kept electronically.

2 Electronic storage should be the norm for those records likely to be kept in the longer term. Documents should be stored in a discrete database available to all JISC staff. Guidance on file formats is available in the DNER Standards and Guidelines. Storage in MS WORD format is likely to be appropriate in the medium term only (up to three years).

3 If records are to be retained or preserved in a usable form, consideration needs to be given to the metadata that is required to ensure continued accessibility and also to demonstrate the authenticity which confers their status as corporate records. Relevant metadata elements are:

- document level
  - document title
  - author/originator
  - date of creation
  - date of last edit
  - version number
  - subject information
  - description or comments
  - document type
  - format
  - e-mail specific attributes (if any)
- record level
  - file/folder classification or directory path
  - linkage between record elements
  - date registered in system
  - audit trail
  - user access restrictions
  - protective marking
  - sensitivity review date
  - presentation versions

.

Access

6 Storage and retrieval are complementary activities which together determine the

sophistication with which the record collections can be accessed and made available for use. Access mechanisms need to:

- prevent time wasted by the user in looking for records
- prevent time or effort wasted in using a non-current version of a record
- provide a common corporate view of the resource
- present a record within the context of a narrative of events to which it relates
- enable the user to search one common access space
- support the segregation and protection of records (eg personal data)
- support comprehensive discovery of relevant records (eg for a DPA or FoI enquiry)

7 One vital feature of electronic records is their ability to open up access to corporate information, providing a richer and more accessible information base for the conduct of JISC business. However, there will still be a need to restrict some forms of access to this material – for example, write access to prevent unauthorized changes being made – and for certain categories of material to restrict all forms of access to defined user groups.

8 Most document management or electronic records management systems will provide for the allocation of rights in an access control table and the attachment to these rights to individual documents or folders. In a less structured environment it will be necessary to rely on less sophisticated facilities, such as the ability to mark documents as read-only.

File Structure and Classification

9 Effective organisation of filing structures for electronic records is achieved by adopting a scheme of naming conventions. These standard rules are designed to be applied to documents, and to electronic folders that contain these documents, in order to enforce consistency in the form of name and in the words and phrases used.

10 A standard for naming documents and folders should be kept simple and clear. It is often preferable to compromise on a broader approach that can be easily understood and remembered, rather than a sophisticated and detailed structure that is less likely to be used in actual practice.

11 Standard terms and forms of name should be used whenever possible. This can apply to:

- names of organizations and people
- names of projects and activities
- logical document types

Full names or acronyms may be used but the practice must be consistent. It is not necessary to repeat information in the document title which is already available from the directory display. For example, the form of a document is held in the file extension and the date of creation/modification is also available.

12 Document titles should contain enough information to identify them if they become detached from the correct folder. Naming conventions should strike the right balance between:

- *brevity* (keeping the title short) and *usability* (usefully describing the content)
- *specificity* (using very precise terms) and *collocation* (grouping under broad headings)

13 Consistent naming rules can link different versions of the same document by including a version number as part of the title. This will also help to provide an audit trail for future tracking of document development. Robust procedures are important for control of document versions in a multi-user environment since there is a danger of inconsistency if a document version is updated separately by different users without coordination.

14 Naming principles should be applied to folders. This may be done by:

- using standard terms for themes and activities which recur across the organisation (for example, project organisation structures)
- using consistent logical labels to describe activities and functions common across the organisation

The use of a thesaurus might be considered in folder titles.

15 In an electronic environment folder structures tend to contain more folders each containing fewer documents than occurs in the paper environment. This may lead to greater depth in the folder structure itself. The length of folder titles can become an issue where a long pathway is built up through the folder hierarchy. In most cases an average of 16 – 20 characters will be adequate, if care is taken to avoid repetition and redundancy.

16 There should be clearly set out rights and responsibilities for folder creation and, where this might be restricted, allocation to specific roles. The role of local staff in maintaining electronic filing structures and the extent to which workgroups are able to create electronic folders themselves should be supported by mechanisms for guiding and controlling the use of terminology.

17 Naming principles should also be applied to web sites. The entry point for project web sites should be contained within the directory containing the project pages. This will enable automated tools to be restricted to the project web site. URLs should:

- be in lower case
- have a consistent file extension, for example all URLs for HTML resources should end *.html*
- omit the file name when referring to the default entry point for a directory, for example use *http://www.foo.ac.uk/about/* not *http://www.foo.ac.uk/about/index.html*
- include the trailing stroke when referring to a directory, for example use *http://www.foo.ac.uk/about/* not *http://www.foo.ac.uk/about*

Sample File Plan

A single filing and records system ensures that people can find the information they need, when they need it. The system should be based on a hierarchy of files/folders:

Level 1 – Functions

Decide the main functions or categories and give each a code. For example:

AC Accounting

HR Human Resources

PM Programme Management

INF Information Policy

Level 2 – Activities

Divide each function/category into activities or sub-categories. For example:

HR/recruitment

HR/pay

HR/training

HR/employee file

Level 3 – Transactions

Activities can be divided into transactions to whatever level is necessary. Two sub-divisions (ie four levels altogether) is the norm. For example:

HR/recruitment/records manager/specification

HR/recruitment/records manager/interviews

HR/recruitment/programme managers

HR/employee file/A N Other

The file plan could use codes or numbers instead of descriptors. For example, the four files at level 3 above might be referenced:

HR 1 / 1 / 1

HR 1 / 1 / 2

HR 1 / 2

HR 2 / A N Other

The complete structure could be represented as follows:

Reference				**Description**
HR	1			Human Resources
	1	1		HR: Policy

	1	1	1	HR: Policy: recruitment
	1	1	2	HR: Policy: training
	1	1	3	etc
	1	1	4	
	1	2		
	1	2	1	
	1	2	2	
	1	3		
	2			
	2	1		
	2	2		
	2	2	1	
	2	2	2	
	2	3		

3.5 Summary Recommendations

Media

- Keep store and access areas free of smoke, dust, dirt and other contaminants.
- Store magnetic media away from strong magnetic fields.
- Transport magnetic media in enclosures with space clearances of 50mm.
- Store in a cool, dry, stable and secure environment
- Acclimatise media before use.
- Use high quality media and devices.
- Keep access devices well maintained and clean.
- Do not place labels on optical disks and/or mark using a pen or pencil. Follow manufacturers recommendations for labelling.
- Minimise handling and use of archival media and/or record number of accesses/use and implement appropriate refreshing.

File Formats

- Use 'open' nonproprietary, well-documented file formats wherever possible.
- Alternatively utilise file formats which are well-developed, have been widely adopted and are 'de facto' standards in the marketplace.
- Identify formats acceptable for the purposes of transfer, storage and distribution to users (these may be distinct).
- Minimise the number of file formats to be managed as far as is feasible/desirable.
- Do not use encryption or compression for archival files if possible.

Media refreshing and reformatting

- Refresh or transfer archive copies to new media at specified times. This should take place:
- within the minimum time specified by the supplier for the media's viability under prevailing environmental conditions;
- when new storage devices are installed;
- when an audit discloses significant temporary or read "errors" in a data resource.
- Employ quality control procedure such as bit/byte or other checksum comparisons with originals to ensure the authenticity and integrity of items after media refreshing.
- Write archive copies with different software to protect data against corruption from malfunctioning or virus- or bug-ridden software.
- Write archive to comparable magnetic media purchased from different suppliers to guard against faults introduced by the media's suppliers into their products or into batches of their products.
- Document actions taken when data resources are copied.
- Retain copies of the digital resource in its original format whenever some information or presentation of the resource may be lost or modified in re-formatting.

Backups and Disaster Recovery Planning

- Create backup copies of data resources at the time of their creation and institute regular backup sessions of data (this may be done automatically over the network)
- Store backup copies on industry standard digital tape or on other approved contemporary media.
- Store backup copies on- and off site. Off-site copies should be stored at a safe distance from on-site copies to ensure they are unaffected by any natural or man-made disaster affecting the on-site copies.

Environmental conditions

- Follow relevant guidance on environmental conditions for storage media
- Prohibit smoking and eating in the storage and work areas.
- Store away from direct sunlight.
- Provide additional protection in the form of enclosures for media.
- Provide storage facilities which minimise the threat from natural disasters such as fire and flood or to magnetic storage media from magnetic fields.
- Ensure any non-digital accompanying materials (e.g. codebooks, operating instructions) are also stored in appropriate environmental conditions.

- Establish guidance and procedures for acclimatising magnetic tape if moving between significant variations in temperature (e.g. tapes moving from very cold external conditions should not be used before being acclimatised to warmer internal conditions).

Audit

- Check media periodically for their readability. Such checking may be conducted automatically in mass storage systems according to parameters set by system operators.
- Check the integrity of data files periodically using checksum procedures. Such procedures may be implemented automatically in mass storage systems according to parameters set by system operators.
- Employ appropriate security systems and procedures to protect the authenticity of items in your holdings (see security below).

Security

- Establish disaster recovery plan.
- Control access to storage facilities and processing areas. Store in separate, preferably lockable area.
- Use passwords and user ids, and other network security procedures.
- Define system and area access privileges for staff.

Assign specific staff responsibilities for data security and storage facilities.

Further information

Beagrie, N. and Jones, M. forthcoming. *Preservation Management of Digital Materials Workbook*. A pre-publication draft is available online.

http://www.jisc.ac.uk/dner/preservation/workbook

AHDS: Guides to Good Practice

http://ahds.ac.uk/public/guides.html

Titles linked at 28 July 2000:

- Archiving Aerial Photography and Remote Sensing Data;
- Excavation and Fieldwork Archiving;
- GIs (Geographic Information Systems);
- Digitising History: a guide to creating digital resources from historical documents;
- Creating Digital Performance Resources;
- Creating and Documenting Electronic Texts;
- Creating Digital Resources for the Visual Arts: Standards and Good Practice;
- Creating a Viable Scholarly Data Resource.

Beagrie, N. & Greenstein, D. (1998a). *A Strategic Policy Framework for Creating and Preserving Digital Collections*. Version 4.0 (Final Draft). ELib Supporting Study p3. Library Information and

Technology Centre, South Bank University, London.

http://ahds.ac.uk/manage/framework.htm

DLM Forum. *Guidelines on Best Practice for Using Electronic Information.* 1997.

http://ispo.cec.be/dlm/documents/guidelines.html

Feeney, M (ed). (1999). *Digital Culture: Maximising the Nation's Investment*. London: The National Preservation Office.

http://www.ukoln.ac.uk/services/elib/papers/other/jisc-npo-dig/intro.html

Lawrence, G.W et al (2000). *Risk Management of Digital Information: A File Format Investigation.* Council on Library and Information Resources. June 2000. (ISBN 1-887334-78-5).

http://www.clir.org.pubs/reports/reports.html

NOF-digitise Technical Standards and Guidelines. Version One; June 2000.

http://www.peoplesnetwork.gov.uk/nof/technicalstandards.html

Preserving Access to Digital Information (PADI).'Storage'.

http://www.nla.gov.au/padi/topics/10.html

'Digital Preservation Strategies'.

http://www.nla.gov.au/padi/topics/18.html

TASI. *Recommendations for Digital Preservation.*

http://www.tasi.ac.uk/building/digprint.html

CAMiLEON (Creative Archiving at Michigan and Leeds; Emulating the Old and the New) Project. Three-year NSF/JISC funded project commenced 1 October 1999. Further details online. Available from the Cedars website:

http://www.www.leeds.ac.uk/cedars/

Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. Rothenberg, Jeff. (1999). Council on Library and Information Resources. January 1999. (ISBN 1-887334-63-7).

http://www.clir.org/pubs/abstract/pub77.html

4. Network Services for Interoperability

4.1 Background to this section

This section lists the protocols and standards that enable content providers to interoperate with other

components in the DNER architecture. These standards are primarily concerned with enabling distributed searching (Z39.50) and the exchange of metadata (Open Archives Initiative, XML/HTTP and RSS). These standards allow presentation (or fusion) services to interact with content providers in order to provide access to the range of material available through the DNER information environment.4.2

4.2 Z39.50

ANSI/NISO Z39.50, along with international counterpart ISO 23950, defines a communications protocol designed to support search and retrieval across distributed, heterogeneous database systems. Whilst Z39.50 does not necessarily offer the enhanced functionality evident in the user interfaces of individual resources within the DNER, it is required if we recognise that users will rarely wish to deal exclusively with a single resource at the start of a search, and will instead desire an ability to cross-search a number of potentially relevant, and possibly very differently structured, sources. Use of Z39.50 as a resource discovery tool in this way need not detract from the ability of product suppliers to offer richly functional interfaces to their resources for those instances in which a user is satisfied to interact with a single resource, or where the nature of the query relies complex additional functionality.

At present, version 3 of this standard forms the basis for multi-database search and retrieval across the DNER. To ensure the maximum degree of interoperability between DNER components, the requirement for adoption of Z39.50 is further qualified by also advocating adherence to the Bath Profile. This Z39.50 profile, developed by the JISC and a number of other international stakeholders, serves to clarify many aspects of the standard, and lays down a minimum set of requirements on systems in order to enable effective inter-working. Z39.50 services within the DNER **MUST** comply with Level of Conformance 0 of Functional Area A or C, depending upon their core audience. Z39.50 services within the DNER **SHOULD** comply with both Functional Areas A and C at Level of Conformance 1.

Access to Z39.50-enabled services, or 'Targets', is by means of a Z39.50 client, or 'Origin'. Traditionally, these Origins have taken the form of proprietary tools from individual system vendors. Increasingly, however, access to Z39.50 Targets is being offered by way of Web-enabled user interfaces that either sit in front of a single Z39.50 Target or manage the submission and return of broadcast searches to one or more remote Targets. Most end users of Z39.50 in the DNER will therefore interact seamlessly with it and other protocols from their Web browser. The emphasis for content providers should be the addition of Z39.50 Target functionality to their products, plus conformance with the requirements of the Bath Profile.

Z39.50
http://lcweb.loc.gov/z3950/agency/document.html

Interoperability Focus
http://www.ukoln.ac.uk/interop-focus/

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

4.3 Open Archives Initiative

The Open Archives Initiative (OAI) is developing a Metadata Harvesting Framework that enables metadata repositories to share their metadata records with other services. In the context of the DNER, it

is anticipated that, where content providers want to share metadata about their own (or other) content with other DNER services, they do this using Open Archives Initiative standards. However, at the time of writing, these standards are under development so it is not possible to be specific about the details of the specifications. Metadata records are likely to be shared using Dublin Core metadata encoded as XML. At this stage, content providers and other DNER services should be aware of the developing OAI standards. A fuller review of OAI will be available as part of the DNER architectural review outlined above.

Open Archives Initiative
http://www.openarchives.org/

4.4 XML/HTTP

In some cases it will be appropriate for content providers and other DNER services to share metadata directly by making it available through their Web servers for collection using HTTP. Where this technique is used it is recommended that the metadata be encoded as XML. In many cases it will be appropriate for the XML-encoded metadata to also conform to the W3C Resource Description Framework (RDF) Model and Syntax recommendation.

Extensible Markup Language
http://www.w3.org/XML/

W3C Metadata Activity
http://www.w3.org/Metadata/

4.4.1 Rich Site Summary

Rich Site Summary (RSS), also referred to as RDF Site Summary, is an XML application for syndicating news feeds. It allows content providers to share metadata about channels of ‘news’ items. Two versions of RSS are available. Version 0.91 is the current standard and the majority of RSS channels currently available conform to this version. A new version is currently under development. Known as RSS 1.0, this version conforms to the W3C RDF Model and Syntax recommendation.

Where content providers or other DNER services make ‘news’ available on their Web site they **SHOULD** also make an RSS channel available. The channel **SHOULD** conform to either the RSS 0.91 or the RSS 1.0 specification.

RSS
http://www.ukoln.ac.uk/metadata/resources/rss/

5. Middleware

5.1 Background to this section

In this document the term ‘middleware’ is used to refer to those services that can be usefully separated from the content providers and the presentation services in order to share them across the DNER as a whole. Four middleware services are currently envisaged, they are:

- Authentication
- Authorisation
- User-profiling

- Collection description

Authentication and authorisation services are currently delivered through the Athens Access Management service. The way in which DNER services should interact with Athens is described below. User-profiling and collection description services are new and will be considered in more detail in the DNER Architectural Study outlined above. However, in is anticipated that collection descriptions should use, or if necessary enhance, the existing RSLP Collection Description Schema. Given that both user-profiling and collection description services can be viewed as being closely related to 'directory' services it may be appropriate to deliver them using a directory technology such as LDAP.

RSLP Collection Description Schema
http://www.ukoln.ac.uk/metadata/rslp/schema/

5.2 LDAP

The Lightweight Directory Access Protocol (LDAP) is a technology that enables distributed access to remote directories of information. Developed initially to provide access to 'whitepages' information (people's names, address, telephone numbers, etc.) held in X.500 directories, LDAP is now more often used to provide access to a diverse range of standalone databases holding information about people, organisations, services, Internet resources, etc.

LDAP may be one of the key technologies that enable DNER services to interact with middleware. LDAP is currently at version 3.

LDAP v 3
http://www.ietf.org/rfc/rfc2251.txt

5.3 Authentication

Authentication by DNER projects and services **SHOULD** be based on the current Athens Access Management service. This includes authentication challenges from content providers to end-users, from presentation services to end-users and between content providers and presentation services or other services within the DNER. It seems likely that, for pragmatic reasons, a mix of username/password based authentication and IP address checking will be required.

Where services need to prompt the end-user for a username/password combination over HTTP (i.e. as part of a Web service), the challenge **SHOULD** be issued using HTTP Basic Authentication.

Athens Access Management
http://www.athens.ac.uk/

Hypertext Transfer Protocol—HTTP/1.0
http://www.w3.org/Protocols/HTTP/1.0/spec.html

5.4 Security and Data Protection

The machines used to deliver DNER services **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.

Services and projects **SHOULD** be managed in accordance with the Information Security Management guidelines laid out in BS7799: Part 1. The management and use of any personal information **MUST** conform to the Data Protection Act 1998.

Where sensitive information is being passed from a client to a server across the network, services and projects **MUST** use Secure Sockets Layer (SSL) to encrypt the data. This includes the transfer of usernames and passwords, credit card details and other personal information. Note that the use of SSL also provides the end-user with an increased level of confidence in the authenticity of the service.

Information Security Management (BS7799)
http://www.bsi.org.uk/bsi/products/msr/bs7799/

Data Protection Act 1998
http://www.hmso.gov.uk/acts/acts1998/19980029.htm

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

6.Presentation

6.1 Background To This Section

This section describes the standards, file formats and guidelines which are applicable to the presentational aspect of a service. It includes file formats for documents, images and multimedia, programming and scripting languages, identifiers and general guidelines on the provision of Web services.

6.2 Access To Resources

Projects **MUST** make information accessible using a Web browser. This will normally be achieved using HTML and the HTTP protocol. If other protocols are used (such as Z39.50) gateways **MUST** be available to provide access to Web browsers. Projects **MUST** be accessible by a variety of browsers, hardware platforms and automated programs.

The Disability Discrimination Act will bring a requirement that all learning (including learning materials) must be accessible to students with disabilities.
In order to conform with this projects **MUST** follow the W3C consortium's Web Accessibility Initiative Guidelines (see http://www.w3.org/WAI/ for Priority 1 and Priority 2.

Projects **SHOULD**, where appropriate, use priority 3 guidelines as well.
The JISC funded Technology Disabilities Information Service can provide help and guidance.

6.3 Document Formats

Projects **MUST** provide HTML for delivery to Web browsers. The HTML **MUST** conform to an open, published DTD such as HTML 4.0 or XHTML.

The appearance of HTML resources **SHOULD** be defined using CSS.

Projects **may** choose to store resources in XML format. XML resources may be transformed to HTML (and CSS) at an application level or by use of XSLT.

The use of proprietary file formats (such as Adobe PDF, Flash, etc.) is undesirable for projects which are intended to be interoperable and to potentially have a long lifetime. However projects **may** choose to use proprietary file formats if alternatives based on open standards are not available or tools to support the open formats are not widely available. Projects which choose to use proprietary file formats **SHOULD** provide strategies for migrating to open formats if and when they become available.

The requirement for use of plugin technologies to provide access to proprietary file formats is discouraged as, even for plugins which are widely deployed, plugins may not be available on platforms such as PDAs, digital TVs, etc. Projects which make use of plugin technologies **SHOULD** be usable if the plugins are not available.

Access by mobile devices **SHOULD** be provided by dynamic generation or conversion from XML to WML or any other appropriate open format supported by mobile devices.

6.3.1 Other Formats

Other file formats may be required for particular purposes: e.g. formats to provide resource metadata, to define news feeds, to support e-commerce and e-business, to implement new technologies such as richer forms of hyperlinking, etc.

Projects **SHOULD** make use of formats defined using XML. For example news feeds **SHOULD** be implemented using RSS or NewsML; enhanced linking technologies **SHOULD** be implemented using XLink and XPointer.

6.4 Graphical Formats

Images **SHOULD** be displayed in GIF (for line-drawings) or JPEG (for photographs). PNG **MAY** be an alternative to GIF.

Video **SHOULD** be displayed using MPEG, Microsoft AVI, ASF or Quicktime formats.

Sound **SHOULD** be delivered using MP3, RealAudio, Microsoft WAV or Sun AU formats.

Note that the use of proprietary file formats (such as Flash) is discouraged, as described above.

The use of emerging new W3C standards such as SMIL (Synchronized Multimedia Integration), SVG (Scalable Vector Graphics) and WebCGM should be considered.

6.5 Programming and Scripting Languages

A number of approaches to server-side scripting may be used, such as using languages such as Perl to provide CGI (Common Gateway Interface) services, server scripting environment such as PHP or ASP, the Java language (and other related components such as JavaBeans), database integration tools or use of content management systems (such as Zope, ColdFusion, etc.) which may have their own scripting environment.

The preferred language for client-side scripting is ECMAscript, which is a standardised version of JavaScript. It should be noted that since users may switch off support for ECMAscript and ECMAscript

may not be supported on platforms such as PDAs, digital TV, etc. project Web sites **SHOULD** be usable if ECMAscript is disabled or not available.

ECMAscript **SHOULD** use the DOM (Document Object Model) to manipulate HTML and CSS elements.

Projects **may** use cookies to maintain state information, but since users may switch off support for cookies and cookies may not be supported on platforms such as PDAs, digital TV, etc. project Web sites **SHOULD** be usable if support for cookies is disabled or not available

SOAP **may** be applicable for providing RPC (Remote Procedure Calls) for Web-based services.

6.6 Identifiers

DNER 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 content provider.

Content providers **MAY** also want to ensure that collections within the resources they are providing are also uniquely and persistently addressable. Due to the limited support for URNs (Uniform Resource Names) and the cost implications of DOIs (Digital Object Identifiers), content providers **SHOULD** make use of the Uniform Resource Identifier (URI) for this purpose. For the moment, the only widely deployed form of URI is the Uniform Resource Locator (URL).

The use of emerging new standards such as DOIs and Urns (for identifying resources) and UDDI (Universal Description, Discovery and Integration) and work on collection descriptions may also be considered. Content providers **MAY** wish to consider assigning a Digital Object Identifier to resources where appropriate.

Projects and services, particularly those concerned with 'reference linking', are encouraged to familiarise themselves with the OpenURL and associated SFX technology.

Uniform Resource Identifiers
http://www.w3.org/Addressing/

Digital Object Identifier
http://www.doi.org/

SFX and OpenURL
http://www.sfxit.com/

6.7 Guidelines For Web Sites

A number of guidelines for providing project Web sites is given below.

6.7.1 Domain Name

If possible try to obtain a domain name for your project Web site, such as www.rdn.ac.uk. If this is not possible try to obtain a domain within your organisation such as rdn.kcl.ac.uk.

You should try to avoid project entry points which have long URLs, such as

http://www.foo.ac.uk/departments/library/projects/jisc/bar.html or entry points which use the tilde convention: http://www.foo.ac.uk/~bar/

Long URLs are difficult to remember and to type, and resources which are located under several sub-directories may not be indexed by search engines. URLs which use the tilde convention may not be regarded as authoritative resources by users accustomed to treating such URLs as personal home pages.

6.7.2 Robots

Project Web sites **SHOULD** contain a robots.txt file. Use of this file is defined by the Robots Exclusion Protocol. It should be used to restrict robots from indexing inappropriate parts of the Web site, such as draft pages, personal pages, etc.

6.7.3 Directory Structure and Naming Conventions

The entry point for project Web sites should be contained within the **directory** containing the project pages. This will enable automated tools (indexed, auditing, mirroring tools, etc.) to be restricted to the project Web site.

URLs **SHOULD** avoid use of unusual characters such as "?" (question mark) when referencing static resources, as such resources are unlikely to be indexed by search engines and are more difficult to cite.

URLs should have a consistent naming convention e.g. all URLs should be in lower case.

URLs should have a consistent file extension e.g. all URLs for HTML resources should end in .html

URLs which refer to the default entry point for a directory **SHOULD** omit the default file name e.g. use http://www.foo.ac.uk/about/ and not http://www.foo.ac.uk/about/index.html

URLs which refer to a directory **should** include the trailing / e.g. use http://www.foo.ac.uk/about/ and not **http://www.foo.ac.uk/about**

6.7.4 Browser Support

Providing Web sites that support open standards and are also usable and fully functional by a wide range of browsers is, unfortunately, not always easy. The browser wars between Netscape and Microsoft have resulted in incompatible approaches and incomplete or ambiguous support for standards.

A number of approaches can be taken to address these difficulties:

- **Stick with mature standards and formats**

  Well-established standards and file formats can be supported and new developments can be ignored. However although this may avoid incompatibility problems, the services are likely to be lacking in functionality, difficult to maintain, and fail to provide user interfaces which are rich, configurable and accessible to a range of users and devices.

- **Deploy the new standards and formats**

  You may choose to deploy new standards and formats in the expectation that the user community will have upgraded to newer browsers by the time the project deliverables are ready. This will allow richer user interfaces to be developed and should allow extra functionality to be provided. However there can be no guarantee that users will, in fact, upgrade their browsers. In addition we

may see a growth of access from non-traditional devices such as mobile phones, digital TVs, games machines, etc. which may not provide the full support for HTML and other standards available on PC systems.

- **Use "safe" features**

  You may choose to use "safe" features if new standards and formats, which will degrade safely with browsers which don't support them and do not cause problems with browsers which have buggy implementations.

- **Use user-agent negotiation**

  You may choose to use user-agent negotiation to check the user's browser for their level of supported features and send file formats and technologies only to browsers which support them.

6.7.5 Web Site Auditing and Evaluation

Projects **SHOULD** audit and evaluate their services to ensure compliance to standards, monitor usage, etc.

6.7.6 Further Information

Further information about the standards and protocols mentioned in this section is given below:

Standards and Protocols

WAI

Specification: http://www.w3.org/TR/WAI-WEBCONTENT/
Further information : http://www.w3.org/WAI/

HTML

Specification: http://www.w3.org/TR/html4/
Further information : http://www.stars.com/Authoring/HTML/

XHTML

Specification: http://www.w3.org/TR/xhtml1/
Further information : http://www.w3.org/MarkUp/Activity

CSS

Specification: http://www.w3.org/TR/REC-CSS2/
Further information : http://www.w3.org/Style/CSS/

XML

Specification: http://www.w3.org/TR/2000/REC-xml-20001006
Further information : http://www.xml.com/pub/

XSLT

Specification: http://www.w3.org/TR/xslt
Further information : http://www.xml.com/pub/2000/08/holman
and http://www.xml.com/pub/2000/08/30/xsltandhtml/

WML

Specifications: http://www.wapforum.com/what/technical.htm
Further information : http://allnetdevices.com/faq/

RSS

Specification: http://my.netscape.com/publish/help/quickstart.html
Further information : http://www.webreference.com/authoring/languages/xml/rss/

NewsML

Specification: Not available
Further information : http://www.iptc.org/newsml.htm

XLink

Specification: http://www.w3.org/TR/xlink/
Further information : http://oasis.oasis-open.org/cover/xll.html

XPointer

Specification: http://www.w3.org/TR/xptr
Further information : http://oasis.oasis-open.org/cover/xll.html

JPEG

Specification: http://www.w3.org/Graphics/JPEG/itu-t81.pdf
Further information : http://www.w3.org/Graphics/JPEG/

PNG

Specification: http://www.w3.org/TR/REC-png.html
Further information : http://www.libpng.org/pub/png/

MPEG

Specification: http://www.cselt.it/mpeg/standards/mpeg-4/mpeg-4.htm
Further information : http://www.cselt.it/mpeg/standards/mpeg-4/mpeg-4.htm

SMIL

Specification: http://www.w3.org/TR/2000/WD-smil20-20000921/
Further information : http://www.w3.org/AudioVideo/

SVG

Specification: http://www.w3.org/TR/2000/CR-SVG-20000802/
Further information : http://www.w3.org/Graphics/SVG/Overview.htm8

ECMAScript

Specification: http://www.ecma.ch/ecma1/stand/ecma-262.htm
Further information : http://webreview.com/pub/98/09/11/feature/sidebar1.html

JAVA

Specification: Not available
Further information : http://java.sun.com/docs/books/tutorial/trailmap.html

Perl

Specification: http://www.perldoc.com/
Further information : http://www.perl.org/

SOAP

Specification: http://www.w3.org/TR/SOAP/
Further information : http://msdn.microsoft.com/msdnmag/issues/0300/soap/soap.asp

DOIs

Specification: http://www.doi.org/handbook_2000/enumeration.html
Further information : http://www.doi.org/handbook_2000/toc.html

Further Reading

W3C
World Wide Web Consortium, which has responsibility for the development of Web standards
http://www.w3c.org/

The Development Of Web Protocols And Formats
Exploit Interactive, issue 1, 10 April 1999
http://www.exploit-lib.org/issue1/web/

Diffuse:
The objective of the Diffuse project is to provide a single, value-added, entry point to up-to-date reference and guidance information on available and emerging standards and specifications that facilitate the electronic exchange of information.
http://www.diffuse.org/diffuse.html

Netscape Search > Computers > Internet > Organizations
http://directory.netscape.com/Computers/Internet/Organizations/Standards

Netscape
http://directory.netscape.com/Computers/Data_Formats/Markup_Languages/HTML/Specifications

Page updated, 12 November, 2001