Irish Manuscripts Commission

Digitisation Policy

45 Merrion Square
Dublin 2
Ireland
www.irishmanuscripts.ie

ISBN 987-1-874280-94-1

Please note hyperlinks in the text will take you directly to many of the reports and websites referenced. Read the document online to enable this functionality. Cite this document as follows: Irish Manuscripts Commission, *Digitisation policy* (Dublin, 2007), 49 pp.

Financial support from the Department of Arts, Sport and Tourism to hold the seminar described in this document is gratefully acknowledged.
The generosity of the Irish Architectural Archive in offering the use of their lecture theatre and conference facilities is also appreciated.

Sincere thanks are due to the members of the Digitisation Task Force, both members from the IMC and the co-opted members from the NAI, NLI, UCD, NUI Galway, NUI Maynooth and PRONI, for their contributions to the production of this report. See page 7 for a full list of Task Force members.

Alastair Dunning (JISC), Niamh Brennan (TCD) and Jennifer Edmond (TCD) have all offered invaluable advice during the editing of this document.
Geneviève Allard (LAC), Dirk Roorda (DANS), Seamus Ross (HATII) and Susan Schreibman (University of Maryland Libraries) commented on a draft and their input is gratefully acknowledged.

Table of Contents

1 Summary and recommendations

1.1 Irish Manuscripts Commission: the need for a policy on digitisation

The Irish Manuscripts Commission (IMC) is a body operating under the auspices of the Cultural Institutions section of the Department of Arts, Sport and Tourism (DAST). Established in 1928, its primary remit is to promote awareness of, and access to, primary source materials for the history, heritage and culture of Ireland. Currently, it achieves this in the following ways:

- editing and bringing to publication historical manuscripts and papers—located in archives and repositories in Ireland and overseas—relating to Ireland that otherwise would not be published;
- its advisory role to the Minister for Arts, Sport and Tourism under Section 48 of the National Cultural Institutions Act, 1997;
- its advisory role to the National Library of Ireland and other archives on the acquisition and publication of historical manuscripts relating to Ireland;
- its liaison with the National Archives Advisory Council (under Section 20 of the National Archives Act, 1986).

The principal activity of the IMC is enabling public and scholarly access and use of manuscript sources through their publication in print form (for details see www.irishmanuscripts.ie). The publication programme is overseen by the board of directors, comprised of policy makers, senior academics, and leaders from the cultural heritage sector. The IMC is therefore ideally positioned as a link between those concerned with the preservation and promotion of Irish cultural heritage and those engaged in research in universities throughout and beyond the island of Ireland.

In the course of reviewing its internal procedures, IMC identified the absence of a policy on digitisation. Simultaneously DAST was examining similar issues in the context of digitisation programmes being funded under the National Development Plan. DAST invited the IMC to consider the issue of digitisation in detail with a view to (a) advising the Minister on best practice in this area and (b) devising a possible framework for developing digital infrastructure in Ireland. In January 2007 this dual requirement was addressed by the formation of a Digitisation Task Force (DTF), under the chairmanship of Professor Jane Ohlmeyer (TCD). IMC members of the DTF included Maurice Bric (UCD), Nicholas Canny (NUI Galway), Chris Flynn (DAST), Thomas O'Connor (NUI Maynooth) and Gerry Slater (PRONI) and the co-opted members included Catriona Crowe (NAI), John Keating (NUI Maynooth), John

McDonough (UCD), Katherine McSharry (NLI), Marie Reddan (NUI Galway), Brian McKenna (NLI) and Larry Murray (PRONI).

The remit of the Digitisation Task Force was:

(1) Draft a policy document on digitisation for the Irish Manuscripts Commission, identifying the key issues in this area;

(2) Organise a high-level seminar on digitisation to take place in 2007;

(3) Bring to the attention of IMC and DAST examples of best practice in place elsewhere;

(4) Advise the Minister for Arts, Sport and Tourism on guidelines for best practice and a possible national digital infrastructure.

This document will fulfil the requirement of (1) and (4) above. The remit of the Digitisation Task Force in (2) and (3) was addressed through the organisation of an international seminar (see Appendix C: Seminar programme for more details) with invited experts from Europe, Canada and North America held on 28 April 2007 at 45 Merrion Square. At the seminar, standards for best practice in digitisation and long-term preservation of digital resources were discussed by leading authorities in these fields in front of an invited audience comprising stakeholders in the digitisation debate (policy makers, commercial service providers, academic researchers, personnel from galleries, archives and libraries). In this way, IMC benefited from expert advice for its own digitisation requirements as well as advice for the possible shape of a digital infrastructure for the national archival and research community, with which IMC is involved.

1.2 Assumptions

While the emphasis of this report is necessarily on the creation of digital objects or resources from older analogue originals, IMC recognises the pressing importance of the need to give consideration at government level to the collection, collation and preservation of digital output that is currently being created (born-digital material).

IMC takes a strong view that digitisation is not a replacement for the practice or funding of primary archive care of collections, but a means for promoting access to and use of archive holdings as well as a means of preservation. Digitised objects not only preserve a sophisticated image of the original object, but they give added value to the collection to which they belong and open it up to new and diverse ways of study and interpretation.

1.3 Recommendations

In preparing its list of recommendations, the DTF has made particular reference to several recent reports on the current situation in Ireland, Europe and North America with regard to digitisation. The reports are represented by the following examples: Higher Education Authority and Forfás, Research

infrastructure in Ireland: Building for tomorrow (Dublin, 2007); Royal Irish Academy, Advancing humanities and social sciences research in Ireland: a report by the Royal Irish Academy (Dublin, 2007); European Strategy Forum on Research Infrastructures, Roadmap for European research infrastructures: Report of the social sciences and humanities roadmap working group, version 4 (Luxembourg, September 2006); DARIAH, A proposal for the Roadmap of the European Strategy Forum on Research Infrastructures (ESFRI); and American Council of Learned Societies, Report of the Commission on Cyberinfrastructure for the Humanities and Social Sciences.

Having reviewed the extensive information available (in print and online) and on foot of discussions with national and international subject-experts, the DTF makes the following **recommendations to the Board of the Irish Manuscripts Commission**:

1. IMC should promote awareness of and co-operation in the area of digital humanities, foster communication between the stakeholders (academic, technical, cultural, commercial partners and policy-makers) and encourage use of international best practice;

2. IMC will explore the practical issues associated with making available in digital format all existing and future IMC publications. Once digitised, this material should be openly accessible and IMC copyright would be protected through appropriate licence agreements;

3. That, subject to a national digitisation strategy and a national digital infrastructure already being in place, IMC should take a proactive role in the digital publication of high-quality historical data sets and primary resources;

4. A national digital infrastructure should be composed of federated local repositories, interlinked and interoperable and meeting Open Archives Information System (OAIS) requirements and linking to a single national trusted digital repository (TDR).

IMC advises the Minister for Arts, Sport and Tourism, as a matter of urgency, that the Irish Government:

1. Develop a comprehensive, inter-departmental framework to implement a national strategy for a digital infrastructure linked to a trusted digital repository;

2. Devise a national strategy that is flexible enough to respond to the dynamic nature of digitisation projects and the technology underpinning them; it must also aim to ensure that digitised resources/ objects are openly accessible, and developed using internationally recognised protocols for data capture and long-term digital preservation;

3. Facilitate the establishment of a national digital infrastructure that is capable of interfacing with the infrastructure currently being developed

for Europe by projects such as DARIAH and DRIVER;

4. Establish a single trusted digital repository for researchers in the humanities and for all users of cultural institution archives;

5. Ensure that adequate leadership and funding is available for setting up a single national trusted digital repository at the heart of a national digital infrastructure;

6. Ensure that the national trusted digital repository maintains a register of all arts, humanities and cultural digitisation initiatives publicly funded in Ireland; follows guidelines for best practice and provides support (advisory, technical) to projects looking to implement them;

7. Ensure capital funding, perhaps as project-linked endowments, to support a digital object throughout its lifecycle, including concept design, management, data capture, user interface tools and functionality, dissemination and long term sustainability.

8. Facilitate the longer term sustainability of national digital assets by requiring applicants for public funding (through HEA, IRCHSS, DAST, etc.), where their project involves representation of analogue materials in digital form (whether through digitisation or text encoding), to include a technical appendix in their research application and to deposit their data according to agreed standards and protocols within the designated national trusted digital repository.

2 Policy development process

2.1 Why now?

A 2007 Forfás report on research infrastructure highlighted Ireland's digital deficit:

> Data archives and repositories remain at an early developmental stage—data access (to existing national and international sources) and data acquisition (storage and updating across the spectrum of disciplines/areas) is weak… The lack of major data repositories and a national depository for qualitative data and repository for research papers is a significant gap… There are important opportunities here for Ireland. Research communities that are currently less well structured than those appearing in the first ESFRI roadmap will be stimulated to participate more effectively in the process from now on. National representatives will be able to participate in discussions about the realisation of these Europe-wide infrastructures. These opportunities will enable Irish policy-makers and researchers not only to engage with nominated ESFRI projects, but also to begin formulating projects of Irish origin in which prominent Irish researchers could play the lead role in future infrastructures.[1]

This digital deficit requires urgent investment if Ireland wishes to benefit from these opportunities as well as comply with stated European objectives. The European Commission in its statement 'i2010: digital libraries' (30 September 2005), set out a vision and strategy for promoting digitisation, fostering online availability, and enabling the long term preservation of the collective memory of Europe.[2] Beyond broader concerns about promoting economic competitiveness and growth, social concerns about public access to culture, monetary concerns about the place of culture in the emerging knowledge economy, and worries related to the place of culture in the emerging intellectual landscape were some of the motivating factors behind 'i2010'. A broad range of Commission-promoted activities from funding programmes to Decisions and Recommendations encouraging action by Member States has resulted from the release of this strategy.

If implemented, the recommendations in this IMC digitisation policy document would help Ireland to contribute to the objectives laid out in the European Commission's Recommendation on the digitisation and online accessibility of cultural material and digital preservation.[3] The timeliness and urgency for taking forward this policy was confirmed on the 22nd of March 2007 when the European Commission took the decision to set up a Member States' Expert

[1] HEA and Forfás, *Research Infrastructure in Ireland: Building for the Future* (Dublin, 2007).
[2] http://ec.europa.eu/information_society/activities/digital_libraries/doc/communication/en_comm_digital_libraries.pdf
[3] (2006/585/EC), 24 August 2006 http://ec.europa.eu/information_society/activities/digital_libraries/doc/recommendation/recommendation/en.pdf

Group on Digitisation and Digital Preservation. Its remit is:

> to monitor progress and assess the impact of the implementation of the Commission Recommendation of 24 August 2006 on the digitisation and online accessibility of cultural material and digital preservation and of the Council Conclusions of 13 November 2006 on the digitisation and online accessibility of cultural material and digital preservation.[4]

2.2 Leadership and co-ordination

To date no single government department or organisation has taken the lead in realising the potential of digitisation as a major resource for increasing access to cultural heritage or for enabling innovation in research and education. This is in spite of the National Representative Group's mission to catalyse action at a National Level in this regard. Responsibility for digitisation is spread across as many as six departments: Taoiseach, Arts, Sport and Tourism, Education and Science, Environment and Local Government, Communications, Marine and Natural resources, and Enterprise, Trade and Employment.

Other European countries, especially the recent accession states, are in a similar position and only now are we beginning to look at possible structures for a national framework for digitisation in the research libraries and archives sector.[5] Thus Ireland is truly in a position to lead the way in terms of putting in place a national framework that co-ordinates existing digitisation services (including digital preservation) and tailors delivery of e-resources (cultural and research) to users' needs. A strategic role could be played by the Secretaries General Group and the Assistant Secretaries General Group looking at e-Government issues in conjunction with the Technical subgroup of the Council of National Cultural Institutions.

Leadership and co-ordination of digitisation effort is essential if the Irish cultural heritage sector is to have a place in the emerging digital landscape, which is fast becoming driven by publishing, entertainment, and broadcasting.[6]

2.3 Fragmentation

At present in Ireland a number of organisations are involved in digitisation projects across a range of disciplines. There is currently no way of identifying the exact number and nature, or even of providing a comprehensive overview, of projects underway or planned by Irish researchers. Potential problems

---

[4] http://ec.europa.eu/information_society/activities/digital_libraries/doc/commission_decision_on_mseg/mseg_en.pdf

[5] See the report commissioned by Joint Information Systems Committee and the Consortium of University Research Libraries: Digitisation in the UK, the case for a UK framework

[6] Screen Digest Ltd, CMS Hasche Sigle, Goldmedia Gmbh, and Rightscom Ltd Interactive content and convergence: Implications for the information society (Luxembourg: DG Information Society and Media, 2006), http://ec.europa.eu/information_society/eeurope/i2010/docs/studies/interactive_content_ec2006_final_report.pdf and the annexes http://ec.europa.eu/information_society/eeurope/i2010/docs/studies/interactive_content_ec2006_annexes.pdf

arising out of this lack of coordination include the inadvertent duplication of research and an inability to ensure that adequate consideration is being given to issues of relevance, quality control and significance. Moreover, the lack of consistency in standards and models for access limits the interoperability between different projects. There is also little progress in terms of certification and ensuring that the authenticity and integrity of digitised and electronic content can be clearly and unequivocally demonstrated.

2.4 Sustainability

The digital environment is a fragile one. Technological, organisational, and market changes all conspire against the long-term viability of digital materials. For example, the rapid pace of technological change threatens the ability to read older file formats, and minute changes or problems with digital bit streams may make the rendering and presentation of a file impossible. To counter these challenges, it is necessary to adopt a curatorial model in the management and preservation of digital data, ensuring that it remains readable and renderable for current and future systems without impacting upon, or prejudicing, the primary source's evidential weight and integrity.[7]

As significant as the technological solutions are to the management of digitised objects, the organisational and institutional infrastructure together with the relationship the user community has with the digital resource are central to the sustainability of such resources. As one expert recently noted '[r]ather than focusing on building collections that merit long-term access at someone's expense, a cultural heritage institution should focus on creating an (online) enterprise that is worth sustaining'.[8] In creating digital resources it is necessary to see the resource as an element in a larger system and to place the overall preservation and curation of the entire collection as the core goal, rather than the sustainability of individual objects within the collection.

2.5 Value for money and improving access

A co-ordinated and strategic approach to digitisation provides a clear context for the work of all institutions considering engaging in and using the results of digitisation projects. The current fragmentation in the system (see 2.3 above) makes it very difficult to ensure that digital projects are achieving value for money, as they cannot avail of a community of practice, economies of scale, or shared standards and channels for dissemination and preservation. Greater co-ordination would allow institutions to pool resources and develop expertise in different areas, avoiding duplication of effort and leading to more efficient delivery of relevant material in digital format.

2.6 Internationalisation of Irish history and culture

The magnitude of these challenges highlights the enormous potential

[7] Extract from Advancing Humanities and Social Sciences Research in Ireland, A Report by the Royal Irish Academy (Dublin, 2007).
[8] Abby Smith. 2003. Issues in sustainability: Creating value for online users. First Monday 8(5) (May). http://www.firstmonday.org/issues/issue8_5/smith/

advantage in working to link data sources together at a European level and to organise national data sources in ways so that they can be linked both nationally and comparatively. [9] At the moment, digitisation projects are not being developed and integrated at the national level, either to ensure the digital resources are interoperable, or in ways that make it possible to coordinate—in terms of data compatibility or project development—with other research projects at the European level.

The recent HEA and Forfás report reiterates the importance of doing this and of enabling Ireland to participate fully in the growing internationalisation of research and resource development of Irish history and culture:

> Selected investments in research infrastructures should be viewed as investments for capacity building in strategic areas of research and career development of young people. International collaboration and researchers' mobility between institutions and participation in international research infrastructure projects provide wider opportunities for researcher training and career development of Irish researchers. High-quality research infrastructures are a necessary prerequisite in attracting foreign researchers and good collaborations in Ireland. Quality infrastructure is also essential in retaining the best Irish research talent in Irish institutions. Investments in infrastructures must be viewed in a national, as well as an institutional context. Thus, issues of access, sharing and transparency, nationally and internationally, will have to be more formally addressed than in the past, including access to existing national databases.[10]

2.7 Key issues

Digitisation is a critical stage in an iterative process. The model proposed by the DARIAH European infrastructure project offers a useful summary of the digitisation continuum or lifecycle:

- Digitise ⇒ Curate ⇒ Preserve
- Discover ⇒ Access ⇒ Deliver
- Connect ⇒ Collaborate ⇒ Use

In terms of identifying the critical aspects concerning the digitisation of documents and images, the creation of data sets and the deployment of digitised resources in the online environment, there are ten key issues that need to be considered: data selection and capture; standards; metadata; storage; access; preservation and curation; digital rights management; strategy; co-ordination; and funding.

[9] This approach will be essential if Ireland's cultural heritage is going to have a place in the European Digital Library as envisaged in European Commission recommendation on the digitisation and online accessibility of cultural material and digital preservation (2006/585/EC), 24 August 2006, http://ec.europa.eu/information_society/activities/digital_libraries/doc/recommendation/recommendation/en.pdf

[10] HEA and Forfás, *Research Infrastructure in Ireland: Building for the Future* (Dublin, 2007).

2.7.1 Data selection and capture

Selecting materials for digital representation and (once an object has been chosen) identifying the most appropriate methods for representing it are crucial processes. Organisation-level digitisation plans need to think beyond current needs and uses towards potential future demands in these areas.

When deciding on the objects to represent digitally, an essential starting point is an understanding of how the end-user will use the resource once it is in digital form. Their anticipated needs will influence how the digital resource is designed, what content is represented and what capabilities to investigate that content are provided (for example, through a user interface tools). Decisions about quality, standards, file format, and end-user functionality all contribute to the overall design of the digital resource and influence the complexities and costs of the digitisation process.

2.7.2 Standards

Observing widely adopted standards in creating metadata plays a critical role in ensuring interoperability with collections on a national and European scale. In the European Commission-funded *Study on the Economic and Technical Evolution of Scientific Publication Markets in Europe*, the authors argued that based on their analysis of the literature and the digital publication landscape itself that 'the use of standards is crucial to enable and facilitate data exchange and communication on the network, and thus definitely improve dissemination and access to scholarly publications.'[11] While agreement regarding appropriate standards is relatively widespread in the digital libraries/ digital humanities community, implementing the same standards between co-operating institutions can be more challenging.

2.7.3 Metadata

International effort in the area of metadata has delivered agreed standards for most document and data types and these are listed below. However, there is a need for training in the use and application of metadata standards and the creation of templates (to extract and validate metadata from HTML resources and MS Office files). Most importantly, there is still a need for the common use of standards between cooperating institutions for both access to and preservation of digital objects.

Proven standards for text-based objects include: the Dublin Core Metadata Initiative (DCMI) for interoperable online metadata standards that support a broad range of purposes and business models; the Text Encoding Initiative (TEI) for international and interdisciplinary standards that enable libraries, museums, publishers, and individual scholars to represent a variety of literary and linguistic texts for online research, teaching, and preservation; the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), an application-independent interoperability framework based on metadata harvesting. Widely accepted internationally, these standards ensure

[11] Mathias Dewatripont *et. al.*, *Study on the Economic and Technical Evolution of Scientific Publication Markets in Europe* (Brussells: DG-Research, European Commission 2006).

repository interoperability and exposure of metadata to web harvesters, crawlers and search engines such as Google.

For non-text based objects common standards include: the Visual Resources Association VRA Core for the description of works of visual culture as well as the images that document them; Categories for Description of Works of Art, which defines categories for use in describing and accessing information about art, architecture or cultural objects, along with any related images; and SEPIADES, SEPIA Data Element Set for cataloguing photographic collections.

2.7.4 Storage

Once digitised or captured, digital content needs to reside somewhere. Local storage can be provided through a combination of online, nearline or offline storage, but a national infrastructure requires a more comprehensive solution. At a national level, an OAIS-compliant repository model would be required for ingest and maintenance of digital content arising from donations from federated, linked local (or institutional) repositories. A national repository would have to meet internationally accepted standards for sustainability, interoperability and trust, i.e. be a Trusted Digital Repository.

2.7.5 Access

Deciding on what elements of the digitisation project can and should be delivered over the Internet depends on the overall objectives of the project: raw data, metadata only, user-interface with tools, or nothing (where the remit is only to preserve the data, rather than to enable public access). Whether access to data will be open or commercial, the cost of each of these options needs to be considered from the outset for each project. Legal issues surrounding copyright and license agreements also play a significant role in decisions regarding the level of access to be provided.

2.7.6 Preservation and curation

Long-term preservation of documents and data is a labour-intensive and currently costly part of the digital object life cycle. The financial commitment will be perpetual. Not only do the data and any associated tools need to be maintained within a well-designed and run Institutional or Local Repository, but in the longer term the digital objects will need to be moved to new technological environments as older systems become obsolete.

While in the short term issues such as security, and adequate backup will need to be addressed, in the longer term decisions will need to taken about the most appropriate approach to preservation for the digital objects (e.g. migration, emulation, regularisation or a combination of these), integration with other digital resources within Ireland and beyond its borders, and how digitisation programmes will respond to the changing needs of researchers.

2.7.7 Digital rights management

The digitisation and retroconversion of manuscripts represents tremendous intellectual effort, comes at a financial cost, and exposes materials to potential misuse or misappropriation. Digital rights management approaches (some technological, some policy-based, and others at institution-level) provide the mechanisms to protect the assets that are created and made accessible through digitisation. Similarly, these approaches can be used to protect assets that are born-digital.

2.7.8 Strategy

Significant questions need to be considered when developing a strategy on digitisation. Any strategy also needs to consider how the data will be accessible to potential users. For instance, will users be permitted to download high-resolution representations of digital objects or will they be limited to low-resolution versions? Will users be able to annotate virtual digital versions with their annotations stored centrally and accessible to other users. Other issues that require consideration relate to how the authenticity of the digital objects will be guaranteed and what kinds of exit strategies will be put in place to ensure that, if a decision were taken not to continue to support the initiative in the longer term, the digital assets that were created by it were not lost.

2.7.9 Co-ordination

Strategic decisions, made at the national level, need to ensure consistency in approach among repositories and adherence to European and international best practice.

Some examples of recent initiatives for local repository-type structures acting in a coordinated and cooperative way include the IUA libraries initiative and the proposed formation of a Digital Humanities Observatory (DHO) (a joint project of the RIA and the Irish universities under the Programme for Research in Third Level Institutions). The latter proposes a structure equivalent to a local repository (see Figure on page 30) and incorporates many of the core features required for a repository operating to best practice standards (recommending standards to its members for digitisation and archiving across a variety of data types, developing access to its resources and providing technical training to its staff).

2.7.10 Funding

Adequate funding will be a critical factor for delivering digitisation projects and for the infrastructure (software, hardware, technical personnel, availability of high quality broadband) required to underpin and sustain these content-based initiatives. In this arena, deciding on what levels of funding can be said to be ‘adequate’ remains challenging. There is a danger, particularly with cultural heritage projects, to fund projects below actual costs.

2.8 Best practice models

On an international scale there are examples that we can point to that demonstrate the viability of such approaches and provide evidence that there is much best practice on which we can rely. Best practice models are available at European (AHDS, DANS) and international level (Library and Archives Canada). Others examples not considered here that have demonstrated that it is viable to produce, distribute and preserve manuscript resources include the National Archives of Scotland and the National Library of New Zealand, and there are many others.

The following summaries briefly describe the approaches of some of these data centres to digitisation of resources in terms of (1) organisation structure (distributed or centralised), (2) funding (central government, portion of income self-generated, partnerships), and (3) activity (capture, delivery and preservation).

2.8.1 Library and Archives Canada model

At Library and Archives Canada (LAC) digitisation is managed by three centres of expertise: the Web Content and Services Division, Sounds and Images Division and the Information Technology Branch.

The Web Content and Services Division leads digitisation from a content perspective, its role is to define institutional priorities in terms of digitisation and to coordinate the efforts of the other centres. The Sounds and Images Division is responsible for the actual digitisation process; these activities are carried out in the Gatineau Preservation Center of LAC. Finally, the Information Technology Branch is responsible for storing and making digital information accessible.

Library and Archives Canada is part of the Department of Canadian Heritage, a department of the federal government of Canada. It has the mandate to make all of its records accessible to all Canadians and therefore, the digitised images of LAC's collection are available on the web free of charge. Funding to digitise LAC collections comes from three main sources: the Department of Canadian Heritage (c. 50% administered via its Canadian Content Online Program); from LAC's own budget (15–25%); and through partnership arrangements (20–25%). LAC has developed partnerships with stakeholders from other government departments (federal, provincial or municipal) as well as profit and non-profit private sector partners.

Digitisation at LAC focuses on access and delivery. All policies, procedures and best practices developed in the last decade have been developed in this context. LAC uses internationally recognised formats and standards, for its imaging, data capture and file storage. While not an official policy of LAC, the issue of digitisation for preservation is, in fact, slowly but surely, becoming a standard practice. With the objective of acquiring the status of a Trusted Digital Repository, LAC staff and stakeholders are at present discussing these very issues.

2.8.2 AHDS model

The Arts and Humanities Data Service (AHDS) is a UK national service aiding the discovery, creation and preservation of digital resources that result from research and teaching in the arts and humanities. Currently, they support preservation of digital data in five subject areas: Archaeology, History, Visual Arts, Literature, Language and Linguistics, and the Performing Arts. A central executive drives the preservation and access policies while expertise in each of the above mentioned areas are based in centres distributed throughout the UK.

The Arts and Humanities Research Council (AHRC) and Joint Information Systems Committee (JISC) jointly fund the AHDS (2006–7 budget c. £1.75 million per annum).

Though their primary brief is preservation of digital data, the AHDS also runs a national training programme and a range of other user-services, promoting awareness amongst a number of UK communities about the importance and value of electronic information.

Most of the data deposited with the AHDS is openly available through their website. The delivery of data varies from collection to collection and is determined by the amount of money available for each project, for example, high-end user-interfaces require money to be constructed in the first instance but also for their maintenance.

The identification and promotion of shared standards is critical to the work of the AHDS. Preserving and exchanging digital information relies upon the widespread adoption of such shared standards, as does a more integrated approach to resource discovery which may help end-users to find the resources they require irrespective of where they are located or how they are stored. The AHDS seeks the widest possible collaboration in identifying such standards and documents and promotes these in its Guides to Good Practice, and its Information Papers, which provide information on a range of issues relating to creating and preserving digital resources.

The AHDS is a partner in the DARIAH initiative of the Social Sciences and Humanities Working Group of the ESFRI.

2.8.3 DANS model

Comparable with the AHDS, the Data Archiving and Networking Services (DANS) is the Dutch national organisation responsible for storing and providing permanent access to research data from the humanities and social sciences.

DANS is an institute of the Royal Netherlands Academy of Arts and Sciences and funded by the Netherlands Organisation for Scientific Research. In terms of its current structure, a central body is responsible for organising the data infrastructure and coordinating the activity of a network of distributed trusted digital repositories. It is staffed by a team of about fifteen people working at the headquarters in The Hague or at one of the networked research centres.

DANS stores and provides access to data, but will also encourage other data centres to do the same, providing they meet certain criteria regarding quality of and permanent accessibility to the data. As more thematic data archives emerge, the emphasis on data access will shift to the network of trusted data repositories. Virtual links between distributed stored data sources—achieved using the Internet and Data Grids—will offer a seamless interface to the user.

Distributed data centres will observe minimum standards (set by DANS) of quality, traceability, accessibility and usability of their data sets. Data sets meeting these standards receive the DANS 'seal of approval'. This compliance is consistent with international standards and guidelines for digital archiving, such as OAIS (the Open Archival Information System[12] developed by NASA), and the Standards for Trusted Digital Repositories developed by the Research Libraries Group (RLG) and the National Archives and Records Administration (NARA) in the United States, and Germany's Network of Expertise in Long-term STOrage of Digital Resources (NESTOR).

2.8.4 European model (DARIAH)

The Digital Research Infrastructure for the Arts and Humanities (DARIAH) initiative is one of several core proposals proposed by the Social Sciences and Humanities Working Group for inclusion in the Roadmap of the European Strategy Forum on Research Infrastructures (ESFRI). DARIAH will provide a coordinated infrastructure for supporting preservation of cultural heritage in Europe and access to research material for the humanities. The key process involved in this will be bringing together the best efforts at national, organisational and individual level in order to launch enhanced Europe-wide actions, initiatives and services. DARIAH will also: assist the development of national services and digitisation programs, particularly in countries where these are non-existent at present; bring together research, education, cultural heritage and 'memory' institutions and organisations in the commercial sector and function to enhance digital scholarship in the humanities and arts across Europe.

The proposed research infrastructure at the core of DARIAH supports and connects the three different areas in the digital lifecycle: researchers, sources and technical infrastructure. DARIAH brings together information-users, information-managers and information-providers and gives them a technical framework that enables enhanced data sharing among research communities. Once DARIAH is in place, digitisation and data dissemination will be supported on a European scale, enabling researchers to engage fully with the widest possible range of information and knowledge.

DARIAH will be based upon an existing network of data centres and services based in Germany (Max Planck Society), France (CNRS), the Netherlands DANS) and the United Kingdom (AHDS). Ireland, Cyprus, Denmark, Greece, Croatia and Slovenia have also been invited to be involved in the preparatory phase.

[12] Reference Model for an Open Archival Information System (OAIS) – ISO 14721 (2002), http://www.ccsds.org/documents/650x0b1.pdf (accessed 10 October 2005).

2.8.5 Lessons learned

All of the above models illustrate the important issues affecting the success of national infrastructures for digitisation:

- Sustainable infrastructure and good strategic decisions are required from the outset at the highest level possible (government);
- A centralised national policy on best practice guidelines for digitising and preserving cultural heritage content and research data resulting from arts and humanities research is essential;
- Any national digitisation strategy needs to have a clear view of the contexts in which the data will be stored, accessed and preserved;
- Advice must be offered for those seeking to undertake digitisation projects, especially if a compulsory technical annexe were to be included in research funding applications;
- Copyright and other legal issues need to be resolved and various usage scenarios considered as an integral part of any policy;
- Core funding for the entire digital life cycle needs to be guaranteed by government if quality digitisation and preservation of cultural and research content is to be sustainable in to the future.

The following table reports some of the current activity at national, EU and international levels in terms of policy documents, repositories and technical standards.

Tabulated list of resources relating to digitisation policy, repositories and standards.

	Policy-making bodies and policy documents	Repositories	Standards/Guides for best practice
IRELAND	Draft DAST policy document in development	No single coordinating body proposed currently	Public Service Standard for Metadata (web administrators) REACH services (open access solution for portal to government services)
	RIA report on 'Research Infrastructures in the Humanities and Social Sciences'	Proposal for a national trusted digital repository = Digital Humanities Observatory	

	Policy-making bodies and policy documents	Repositories	Standards/Guides for best practice
	Irish Universities and Higher Education Institutions	Federated institutional repositories with national harvesting service and discovery portal: Three-year IUA Library sectoral project funded under HEA's Structural Innovation Fund. Partners: Irish Universities. Project starts April 2007	OAI-PMH Core set of metadata for harvesting from Irish federated repositories (under development)
		National Digital Learning Repository (NDLR) [All Irish universities & IoTs]	IEEE LOM / NDLR LOM, SCORM
		Irish Virtual Research Library and Archive Project creating digitised humanities material and researching digital repository development and deployment (UCD).	IVRLA Digitisation Processes, OAI-PMH, IVRLA MODS Implementation Guidelines, IVRLA Regulatory Issues
		TARA (Trinity's Access to Research Archive) *[TCD]*	Dublin Core (Qualified), VRAcore, OAI-PMH
		DCU Institutional Repository	Dublin Core, OAI-PMH
		NUI Maynooth ePrints and eTheses Archive	Dublin Core, OAI-PMH
		Transfer of Expertise in Technologies of Editing (TEXTE) *project at NUI Galway*	
		CELT: Corpus of Electronic Texts *[UCC]*	TEI
		Irish Social Science Data Archive (ISSDA) (operates within the Geary Institute, UCD; managed jointly by UCD, Economic and Social Research Institute with the support of the Central Statistics Office.	
EUROPEAN UNION	ESFRI Roadmap for Research Infrastructures in the Humanities and Social Sciences.	DARIAH (centralised group managing preservation standards and offering support and advice to distributed TDRs).	AHDS Good Practice Guides.
	Dynamic Action Plan (National Representatives Group of Member States)		MINERVA List of Good Practice Guides (European and international).

	Policy-making bodies and policy documents	Repositories	Standards/Guides for best practice
	i2010 Digital Libraries (European Commission)		MINERVA List of Good Practice Guides (European and international).
INTERNATIO-NAL	American Council of Learned Societies		Report of the Commission on Cyberinfrastructure for the Humanities and Social Sciences
	Research Libraries Group		Trusted Digital Repositories: Attributes and Responsibilities
			Open Archives Information System (OAIS) Dublin Core Metadata Initiative

3 Towards a policy on digitisation

3.1 The IMC perspective

The primary goal of IMC is to ensure continued access to the manuscript heritage of Ireland and this will be done using the most appropriate methods whether this is print or digital or a combination of both. The DTF sees digitisation as a way to maximise the unexplored potential of IMC publications by facilitating a greater volume of remote access, and by providing the enhanced functionality (for example, advanced text and image searching) that is intrinsic to digitised data. In addition to preserving our publications for future generations, digitisation of IMC resources if carried out in conjunction with other institutional collections would also reflect national and international priorities for wider access to, and enhanced use of, cultural materials that have a combined value greater than their component parts. For example it makes possible collaborative scholarship across Ireland and at an international level and it contributes to moving humanities scholarship into the age of eScience.

One of the core principles of IMC's policy on digitisation is that it must represent value for money and eliminate duplication of effort. It is widely acknowledged that a long-term commitment from government is required in terms of funding the capture and long-term preservation of digital data, especially in terms of the qualified technical staff required to support digitisation activities. However, offsetting this cost somewhat is the fact that much of the information and software required to achieve this is available for free, and the up-skilling of the research and support infrastructure that will be required to take advantage of these tools will have a long-term benefit for the Irish society and economy. Adopting existing European and international best-practice guidelines to form an Irish (national) digital infrastructure model represents a cost-effective way for a variety of national organisations to participate in sustainable, long-term preservation of digital research resources and items of cultural heritage.

The DTF regards inter-institutional cooperation on digitisation policy as critical and the framing of a top-down national strategy on digitisation as urgent. For example, see the recommendations of the *Dynamic Action Plan for the EU co-ordination of digitisation of cultural and scientific content* produced by the National Representative Group of the Member States and the associated work that MINERVA has done on digitisation policies, and the Social Sciences and Humanities Roadmap Working Group report to the European Strategy Forum on Research Infrastructures. Both reports emphasise that a government-level strategic policy should be in place from the outset rather than a piecemeal, reactive measure.

A national policy (possibly coordinated at the outset by the eGovernment

Working Group of the Secretaries General Group and thereafter by the Council of National Cultural Institutions and the Higher Education Authority) should correlate and harmonise activities carried out in digitising cultural and research content with a view towards the creation of data that are compatible with a common European platform (see above), protocols for digitisation, metadata, long-term accessibility and preservation. With this in mind, the IMC would anticipate depositing copies of its publications in a suitable national trusted digital repository, rather than undertaking management of long-term preservation itself.

This document focuses on recommendations for best practice for the IMC in terms of preserving digital copies of its publications into the future and increasing access to the information contained within them. It does not attempt to re-invent the wheel by drafting detailed technical guidelines for best practice in digitisation and preservation of digital resources. These guidelines are available online from a variety of sources, for example, see Digitisation guidelines: a selected list (MINERVA), the NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials, the guidelines and protocols of the Arts and Humanities Data Service and the British Library Preservation Management of Digital Materials Handbook for the creation and use of digital resources.[13]

3.2 Who is the IMC's audience?

Informing the evaluation of what to digitise should be the identification of the audience for the digitised data and how they might use it, now and in the future. The profile of the audience for IMC publications is (1) academic and scholarly users (historians, literary scholars, linguists, historical geographers), (2) genealogists and (3) readers and life long learners with a general interest in history, including local history societies and (4) policy makers and funding agencies that support the humanities. These users require superior searching tools and detailed indexes. If IMC publications were to be made available online, users would require well designed, front-end (user interfaces and searching mechanisms) and back-end (tools, software and hardware infrastructures) delivery of digital resources.

By making the material accessible in this way new audiences for IMC publications will emerge and new ways of working among the IMC audience will evolve. In considering this, it is important to think outside traditional constituencies because digital representation changes the reach of information (and how it can be used) so dramatically as to engage new communities of users.

3.3 What to digitise?

IMC should explore the possibility of making available in digital format all existing and future IMC publications, on an open-access basis. IMC copyright would be protected through proper license agreements;

[13] See Appendix A: Further reading for a list of documents dealing with digitisation guidelines and protocols.

Subject to a national strategy and national digital infrastructure being in place, the IMC would be in a position to take a proactive role in the digital publication of high quality historical data sets and primary materials (archival, printed, audio-visual etc.).

3.4 Sustainability

Preservation and maintenance of digitised data is a labour-intensive task and needs considerable resources and support if we are to ensure that the digital resources we create now are available for future researchers. **At this time, IMC does not envisage managing digital preservation of its publications itself**. Rather it would ideally feed correctly formatted digital data into a national trusted digital repository where the data would then be preserved.[14]

As the work of researchers provides IMC with many of its publications, IMC recommends that all public funding bodies (such as the HEA, IRCHSS and DAST) **include a requirement for a technical appendix in all relevant applications for research funding** (see Appendix B for a copy of the technical appendix used by the Arts and Humanities Research Council (AHRC) in the UK, and see also the Institute of Museum and Library Services' National Leadership Grants application form at http://www.imls.gov/applicants/grants/pdf/NLG_2007.pdf). This would have the effect of harmonising the approach of individual researchers and institutions to digitisation and depositing digital data with trusted digital repositories. A suitable, independent (external) review body should be involved in the evaluation of these annexes in order to identify the research groups that are following internationally recognised protocols and addressing fundamental issues regarding the preservation of, and access to, the data and resources they generate during the course of their research, as well as issues such as project management and dissemination that will guarantee that the digital resources do come to the widest possible audience and deliver value for public funding.

3.5 Proposal for a best practice model for Ireland

We propose the following as a template for a national digital infrastructure for Ireland:

- A single trusted digital repository called, for example, the Irish National Data Centre, would serve the digital preservation needs of all of

[14] A carefully negotiated Service Level Agreement would underpin and regulate this arrangement.

Ireland. A government-level national body would set (and update) the standards and best practice guidelines for digitisation and preservation for the Irish National Data Centre with due reference to current European and international standards and protocols.

- This national trusted digital repository would be responsible for the long-term preservation of digitised objects deposited with it. It would ensure its content was backed-up to a mirror site or a combination of a mirror site and a dark archive.

- Local or institutional repositories would be linked to the Irish National Data Centre and deposit data with it, benefiting in turn from federated search services across all linked repositories. The PRTLI-funded IVRLA project is deploying a repository infrastructure for digital content, which might serve as a model for other local repository structures.

- The Irish National Data Centre would link Irish digitised objects (from cultural and research bodies) into existing European and international repositories and networks such as DARIAH and DRIVER. This would enhance and augment international access to Irish cultural and heritage objects.

- Local repositories representing research groups, projects or data centres creating and managing digitised data would do so in an agreed format. Staff at local repositories can augment their expertise with cross-discipline training, for example, in arts/humanities and computing.

- Local repositories feed their standardised data into the Irish National Data Centre where license agreements would determine the use of deposited content. All depositors have the option of using the Irish National Data Centre as their sole means of digital preservation thereby saving on money and duplication of effort.

- The Irish National Data Centre will produce annual information and audit reports of their preservation and curation activity, which can be used to assure the trust of the user communities depositing data with it.

It is important when considering a best practice model for Ireland to remember that the process is an iterative one requiring in-built flexibility to respond dynamically to the changes and developments in digitisation and digital preservation. We believe the following actions are important for success:

- The development of a coordinated and comprehensive government roadmap on digitisation.

- A programme for increasing access to cultural content of museums, libraries, archives and galleries and the research emanating from Irish universities using the most up-to-date technologies.

- The creation of a national body to oversee and manage the Irish National Data Centre (a Trusted Digital Repository). A technical subgroup of this body would set guidelines for best practice in digitisation and preservation of all digital resources and keep these up-to-date through contact with EU and international bodies setting the standards.

- The national body in charge of the Irish National Data Centre would maintain a register of digitisation programmes ongoing in Irish universities, libraries and cultural bodies as part of its user interface to newly curated accessions.

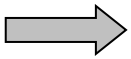
- Financial support for the Irish National Data Centre should come from central government; this will ensure consistency and continuity of digital preservation for Irish cultural heritage into the future.

- As far as possible, user interfaces should enable easy and open access to, and detailed usage of, all digital collections preserved by the Irish National Data Centre on behalf of distributed local repositories, including the national cultural institutions. This will ensure the widest possible dissemination of Irish cultural heritage within Ireland and beyond these shores.

The following schematic diagram represents the components of a national digital infrastructure and how they might interoperate.

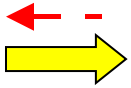
KEY TO DATA EXCHANGES AND FUNCTION OF COMPONENTS IN SCHEMATIC DIAGRAM

Local Repository	Irish National Data Centre (a Trusted Digital Repository)
Deposits digitised objects with the Irish National Data Centre.	Preserves digital objects deposited by Local Repository (LR), to archival quality, for the long term using a mirror site and a dark archive.
Provides user interfaces locally to its own digitised objects (browse-quality text, image and audio-visual files), but adds search/display tools that show similar/related digital objects available through Irish National Data Centre, i.e. available in collections of other LRs.	Curates its content (active preservation) and highlights newly added data available through its user interface. This could include lists of projects with a digitisation component approved by research funding bodies, where these bodies act as LRs and deposit digital objects (approved project-proposals) with the Irish National Data Centre.
Ensures **compliance** with technical and operational requirements of data deposition with the Irish National Data Centre by using their stipulated formats and metadata.	**Recommends** formats for metadata to be used by LRs depositing digitised objects with Irish National Data Centre.
Draws-up license agreements to address **copyright** concerns and access to its digital objects via Irish National Data Centre.	Carries out **validation and authentication** of objects it receives. Implements terms of license agreements with LR when providing access to data through its user interface(s).
Can, depending on institutional policy, use Irish National Data Centre as its sole option for long-term **preservation** and management of its digital resources; this allows for savings in its annual budget and reduces duplication of preservation effort	Performs long-term **preservation** function for Local Repository thereby saving the LR the cost and administration of implementing a preservation policy.



Digital preservation: secure transfer of stored data from the Irish National Data Centre to back-up (mirror) site and/or a dark archive.

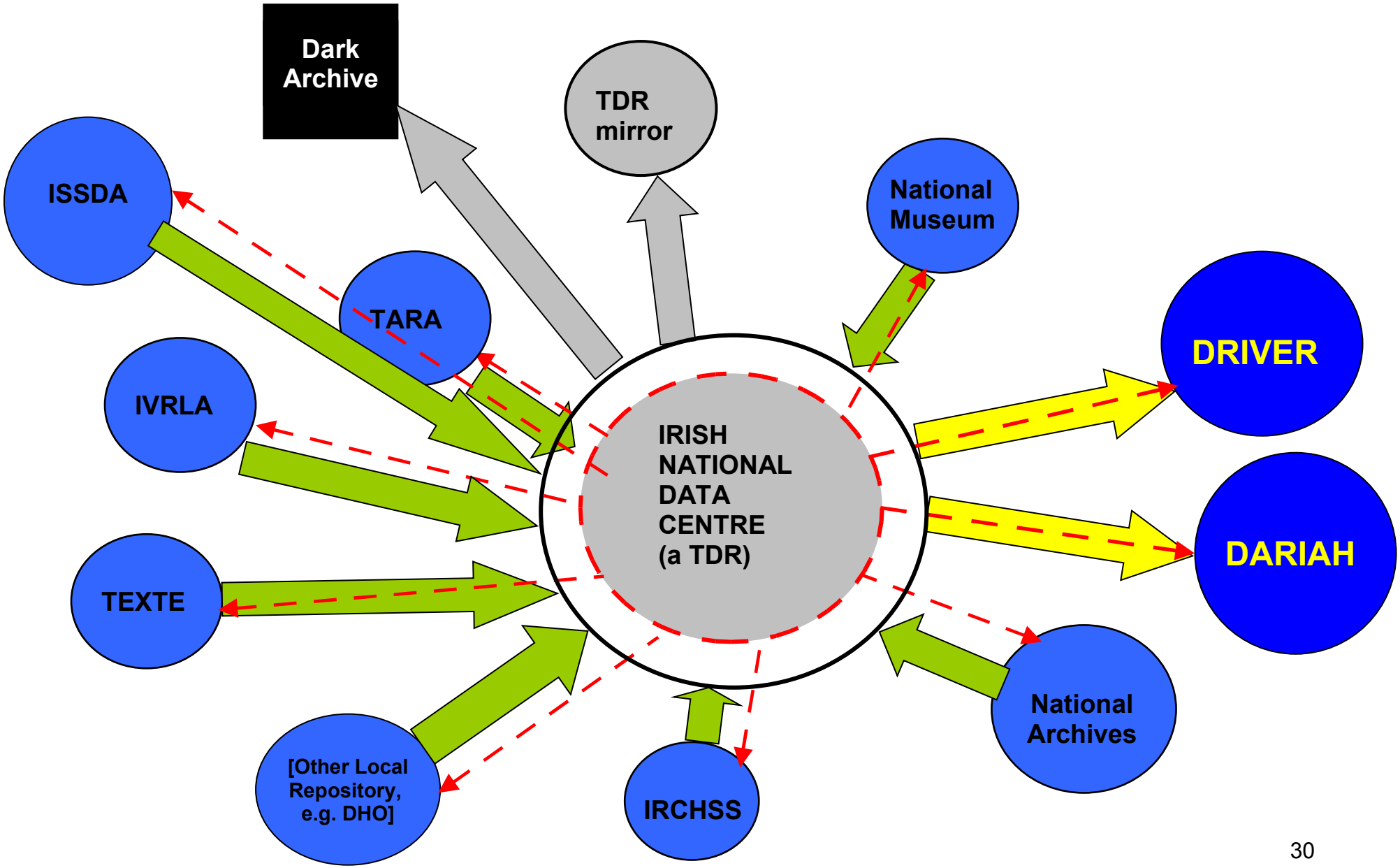
Donation of digitised objects from Local Repository to the Irish National Data Centre.



Access to digitised objects from the Irish National Data Centre collection being viewed at interface(s) in Local Repositories and through links to European projects.

For explanations of dark archive, mirror site, TEXTE, TARA, IVRLA, IRCHSS, DRIVER and DARIAH, see Appendix D: Glossary.

Proposal for a national data centre at the heart of a national digital infrastructure.



Appendix A: Further reading

Overview of digital infrastructure literature

National

The Department of Arts, Sport and Tourism (DAST) proposes to adapt this digitisation policy document, pending Ministerial approval. In its final form, it may suggest the formation of a National Cultural Digitisation Group (NCDG) to oversee and steer a national digitisation programme, possibly as part of the Council of National Cultural Institutions.

The Royal Irish Academy (RIA) is an all-Ireland, independent academic body that promotes study and excellence in the sciences, humanities and social sciences. It has recently published a report entitled *Advancing humanities and social sciences research in Ireland*. Among other things, the report recommends: (1) the development of a national register of research and a repository for published research papers and address the lack of major data repositories for Humanities and Social Sciences (HSS) research; (2) the establishment of a Digitisation Stakeholders Forum to provide a direct mechanism for the input of all relevant stakeholders in the development and implementation of a national digitisation policy. The RIA envisages that such a forum might inform the establishment of a national policy on digitisation for the HSS and should also make proposals for curation and preservation, ensuring that data remains readable and compatible with current and future technical systems, and that common standards and formats are adopted in line with international best-practice.

The Irish Virtual Research Library & Archive (IVRLA) is a major digitisation and digital object management project launched in UCD in January 2005. Conceived as a means to preserve elements of UCD's main repositories and increase and facilitate access to this material through the adoption of digitisation technologies, the project will also undertake research into the area of interacting with, and enhancing the use of, digital objects in a research environment through the development of a digital repository. When fully implemented, the IVRLA will be one of the first comprehensive digital primary source repositories in Ireland. The project is a component of the Humanities Institute of Ireland (HII) and is funded by the Programme for Research in Third Level Institutions (PRTLI) Phase 3.

The Irish Universities Association (IUA) Library group has initiated the construction of a national network of federated open-access institutional repositories along with a national harvesting service and discovery portal. The models for this network are DAREnet (Netherlands) and ARROW (Australia). This is a three-year IUA sectoral project funded under HEA's Structural Innovation Fund that started in April 2007. The project partners are the seven Irish universities. The aim of the project is to significantly increase access to university research publications and postgraduate theses in order to showcase Irish research output and maximise its impact on the international stage. The project website is hosted by HEAnet and is available at: http://www.learningcontent.edu.ie/iual-wg/

Individual open access repositories established prior to the IUA SIF-funded project described above, which will become part of that project include:

TARA (Trinity's Access to Research Archive) (http://www.tara.tcd.ie/)
DCU Institutional Repository (http://eprints.dcu.ie/)
NUI Maynooth ePrints and eTheses Archive (http://eprints.may.ie/)
Irish Social Science Data Archive (http://www.ucd.ie/issda/)
Corpus of Electronic Texts (CELT), based in UCC, aims to 'to bring the wealth of Irish literary and historical culture (in Irish, Latin, Anglo-Norman French, and English) to the Internet in a rigorously scholarly and user-friendly project for the

widest possible range of readers and researchers.'

The recently commissioned National Digital Learning Repository (NDLR) is a HEA-funded pilot project between Irish Universities and the Institutes of Technology to support the collaboration and sharing of learning and teaching resources through a trusted digital repository. The project is directed by TCD and the repository is hosted by HEAnet.

Transfer of Expertise in Technologies of Editing (TEXTE) at NUI Galway is a Marie Curie Transfer of Knowledge project. The project will fund six post-doctoral European researchers to work at NUI, Galway, creating electronic archives and editions of historical and literary texts using new technologies of imaging, text-encoding, editing, and hypermedia publication.

The Irish Public Service Metadata Standard offers guidelines for managers and those with the task of creating the metadata associated with online public service web resources.

Through the Reach project, the government of Ireland has developed a portal (the Public Services Broker) to enable interaction between customers and government departments with a public service interface. The project was mandated to develop and implement an integrated set of processes, systems and procedures to provide a standard means of access to public services. The specifications of the multi-channel architecture and functionality for the Public Services Broker are openly accessible.

European

In 2005, Joint Information Systems Committee (JISC) and the Consortium of University Research Libraries (CURL) commissioned a review of the state of digitisation in the UK. The resulting report, *Digitisation in the UK: the case for a UK framework* (2005), makes the case for strengthened co-ordination and the establishment of a UK framework to ensure future digitisation projects are better executed, more sustainable, and respond directly to the needs of the research community.

The European Strategy Forum on Research Infrastructures (ESFRI) is the umbrella body coordinating the development of a plan for the construction of large-scale research infrastructure to underpin innovation and progress across several disciplines. It has recently produced its final recommendations in a report entitled European roadmap for research infrastructures, Report 2006.

The Social Sciences and Humanities Working Group of the ESFRI has also produced a detailed report of its recommendations for six projects/proposals supporting the goals of the 'roadmap' in its document Roadmap for European research infrastructures: Report of the social sciences and humanities roadmap working group, version 4 (Luxembourg, September 2006).

Digital Research Infrastructure for the Arts and Humanities (DARIAH) is one of six approved projects of the Social Sciences and Humanities working group of the ESFRI. The project is described in detail on the DARIAH website, A proposal for the Roadmap of the European Strategy Forum on Research Infrastructures (ESFRI). Four European data centres (AHDS (UK), CNRS (FR), DANS (NL) and the Max Planck Institute (D)) have drawn up a plan for a digital infrastructure that supports research in the Arts and Humanities and benefits from existing infrastructure such as European Cultural Heritage Online (ECHO) open access project. All local digitisation projects within national domains not currently involved in DARIAH are welcome to join the centrally-coordinated research infrastructure. DARIAH has an estimated construction cost of €10 million and an annual operational cost of €4 million.

MINERVA is a thematic network of Member States' Ministries to discuss, correlate and harmonise activities carried out in the digitisation of cultural and scientific content. The MINERVA website has an excellent list of guides to good practice in digitisation.

Strategies for a European Area of Digital Cultural Resources: Towards a Contiuum of Digital

Heritage. Conference report, The Hague, The Netherlands, 15-16 September 2004.

International

Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences. The term cyberinfrastructure was coined by the US National Science Foundation to describe the new research environments in which high-performance computing tools are available to researchers in a shared network environment. The ACLS feels it is important for scholars in the humanities and social sciences to participate in the design and construction of these tools and environments.

Research Libraries Group (RLG) and Online Computer Library Center (OCLC), *Trusted Digital Repositories: Attributes and Responsibilities*, An RLG-OCLC report (Mountainview, CA, 2002).

Susan Schreibman, Ray Siemens, John Unsworth (eds), *A Companion to Digital Humanities* (Oxford: Blackwell, 2004). http://www.digitalhumanities.org/companion/

Reference works on digitisation

Several lists of reading material on digitisation exist on the Internet. The material recommended here is compiled from two main sources: a publication of the UK Office for Library Networking (UKOLN) and the EU thematic network MINERVA section on guides to good practice in digitisation; it is not exhaustive by any means.

General

Canadian Library and Archives Best Practice Manual. The purpose of this document is to provide publishers with an introduction to the pros and cons of various approaches to electronic publishing. Through the use of hyperlinks, it highlights examples of some of the most creative and efficient uses of new media technology for publishing, and points to sites maintained by organisations that provide resources for online publishers.

Howard Besser and Jennifer Trant, *Introduction to imaging: issues in constructing an image database*. Santa Monica, Calif.: Getty Art History Information Program, 1995.

Anne R. Kenney and Stephen Chapman, *Digital imaging for libraries and archives*. Ithaca, N.Y.: Cornell University Library, Department of Preservation and Conservation, June 1996.

Hartmut Weber and Marianne Dörr, *Digitisation as a method of preservation? Final report of a working group of the Deutsche Forchungsgemeinschaft (German Research Association)*. Amsterdam: European Commission on Preservation and Access, 1997. http://www.clir.org/pubs/abstract/pub69.html

Susan Jephcott, 'Why digitise? Principles in planning and managing a successful digitisation project.' *New Review of Academic Librarianship*, 4, 1998, pp. 39-52.

Michael Day, Preservation of electronic information: a bibliography. Bath: UKOLN, UK Office for Library and Information Networking. http://homes.ukoln.ac.uk/~lismd/preservation.html

Guidelines for digital imaging: papers given at the Joint National Preservation Office and Research Libraries Group Preservation Conference in Warwick, 28th - 30th September 1998. London: National Preservation Office, 1998. http://www.rlg.org/preserv/joint/

Anne R. Kenney, 'Mainstreaming digitisation into the mission of cultural repositories.' In: *Collections, content, and the Web*. Washington, D.C.: Council on Library and Information Resources, 2000, pp. 4-17. http://www.clir.org/pubs/abstract/pub88abst.html

Anne R. Kenney and Oya Y. Rieger, *Moving theory into practice: digital imaging for libraries and archives*. Mountain View, Calif.: Research Libraries Group, 2000.

Selecting material for digitisation

NISO Framework Advisory Group, A framework of Guidance for building good digital collections (Bethseda, MD, NISO Press, 2004).
http://www.niso.org/framework/framework2.pdf

Ricky Erway, ed., *Selecting library and archive collections for digital reformatting: proceedings from an RLG Symposium held November 5-6, 1995 in Washington, DC*. Mountain View, Calif.: Research Libraries Group, 1996.

Dan Hazen, Jeffrey Horrell, Jan Merrill-Oldham, *Selecting research collections for digitization*. Washington, D.C.: Council on Library and Information Resources, 1998.
http://www.clir.org/pubs/abstract/pub74.html

Seamus Ross, 'Strategies for selecting resources for digitization: Source-Orientated, User-Driven, Asset-Aware Model (SOUDAAM).' In: Terry Coppock, ed., *Making information available in digital format: perspectives from practitioners*. Edinburgh: The Stationery Office, 1999, pp. 5-27.

Abby Smith. *Strategies for Building Digitized Collections. Council on Library and Information Resources* (CLIR), Washington, D.C., 2001
http://www.clir.org/pubs/reports/pub101/contents.html

Digital representation

Arms, Caroline R., and Carl Fleischhaur. *Sustainability of Digital Formats: Planning for Library of Congress Collections*. Washington, D.C. : Library of Congress, http://www.digitalpreservation.gov/formats/sustain/sustain.shtml#adoption.

Clausen, Lars R., main author. *Handling File Formats*. Denmark: State and University Library, Arhus, and the Royal Library, Copenhagen, May 2004.
http://netarchive.dk/publikationer/FileFormats-2004.pdf#search=%22handling%20file%20formats%22

Technical Advisory Service for Images (TASI). Funded by JISC, TASI manages a growing compendium of advice and guidance in the area of digital representation.
http://www.tasi.ac.uk/

Andrew Wilson et. al., *Moving Images and Sound Archiving Study*. Arts and Humanities Data Service, United Kingdom : Final Draft, June 2006, p. 29.
http://www.jisc.ac.uk/uploaded_documents/Moving%20Images%20and%20Sound%20Archiving%20Study1.doc

Textual resources

Peter Robinson, *The digitization of primary textual resources*. Oxford: Office for Humanities Communication, 1993.

Peter Robinson, *The transcription of primary textual sources using SGML*. Oxford: Office for Humanities Communication, 1994.

Guides to good practice

HATII at the University of Glasgow and the National Initiative for a Networked Cultural Heritage (NINCH), 2002, *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials* (Washington, D.C., 2002).

http://www.ninch.org/guide.pdf

North Carolina ECHO (Exploring Cultural heritage Online), *Digitisation Guidelines* (NCECHO, 2004). http://www.ncecho.org/Guide/2004/toc.htm

Arts and Humanities Data Service, *Creating a viable scholarly data resource*. London: AHDS, ca. 1998. http://ahds.ac.uk/deposit/viable.html

Sean Townsend, Cressida Chappell and Oscar Struijvé, *Digitising history: a guide to creating digital resources from historical documents*. Oxford: Oxbow Books, 1999. http://hds.essex.ac.uk/g2gp/digitising_history/index.html

Alan Morrison, Michael Popham and Karen Wikander, *Creating and documenting electronic texts: a guide to good practice*. Oxford: Oxbow Books, 2000. http://ota.ahds.ac.uk/documents/creating/

Catherine Grout, *et al*., *Creating digital resources for the visual arts: standards and good practice*. Oxford: Oxbow Books, 2000 (forthcoming). http://vads.ahds.ac.uk/guides/creating_guide/contents.html

Metadata

Michael Day, 'Metadata', in Ross, S. (ed.), DCC Digital Curation Manual (Glasgow, 2005) http://www.dcc.ac.uk/resource/curation-manual/chapters/metadata.

RLG Working Group on Preservation Issues of Metadata, *Final report*. Mountain View, Calif.: Research Libraries Group, 1998. http://www.rlg.org/preserv/presmeta.html

PREMIS Working Group, 2005, *Data Dictionary for Preservation Metadata*, Dublin, OH and Mountain View, CA: OCLC and RLG, *http://www.oclc.org/research/projects/pmwg/premis-final.pdf*. *Note* PREMIS provides a core set of preservation metadata elements along with a detailed description in the form of a Data Dictionary. While PREMIS is weak on technical metadata (they acknowledge this), in general the PREMIS data dictionary is very powerful.

Bernard J. Hurley, John Price-Wilkin, Merrilee Proffitt and Howard Besser, *The Making of America II Testbed Project: a digital library service model*. Washington, D.C.: Council on Library and Information Resources, 1999. http://www.clir.org/pubs/abstract/pub87abst.html

The Text Encoding Initiative (TEI) Guidelines are an international and interdisciplinary standard that enables libraries, museums, publishers, and individual scholars to represent a variety of literary and linguistic texts for online research, teaching, and preservation. The TEI standard is maintained by a Consortium of leading institutions and projects worldwide.

An American initiative, Cataloging Cultural Objects (CCO) is a data content standards initiative for the cultural heritage community. CCO web resources include cataloging examples, training tools and presentations for use by practitioners, excerpts from the CCO print publication, etc. Sponsored by the Visual Resources Association, CCO activities center on educational efforts to promote widespread acceptance of cataloging best practices for the community.

See also the list of commonly-used metadata formats in the Arts and Humanities on the AHDS website at http://ahds.ac.uk/metadata/arts-humanities-metadata-formats.htm

Digital preservation

General

Don Waters and John Garrett, eds., *Preserving digital information, report of the Task Force*

on Archiving of Digital Information. Washington, D.C.: Commission on Preservation and Access, 1996. http://www.rlg.org/ArchTF/

Margaret Hedstrom, 'Digital preservation: a time bomb for digital libraries.' *Computers and the Humanities*, 31 (3), 1997, pp. 189-202.

Deanna Marcum, 'A moral and legal obligation: preservation in the digital age.' *International Information & Library Review*, 29 (3/4), 1997, pp. 357-365.

Neil Beagrie and Daniel Greenstein, *A strategic policy framework for creating and preserving digital collections*. London: South Bank University, Library Information Technology Centre, 1998. http://ahds.ac.uk/manage/framework.htm

Margaret Hedstrom and Sheon Montgomery, *Digital preservation needs and requirements in RLG member institutions: a study commissioned by the Research Libraries Group*. Mountain View, Calif.: Research Libraries Group, December 1998.
http://www.rlg.org/preserv/digpres.html

Jeff Rothenberg, *Avoiding technological quicksand: finding a viable technical foundation for digital preservation*. Washington, D.C.: Council on Library and Information Resources, 1999.
http://www.clir.org/pubs/abstract/pub77.html

David Bearman, 'Reality and chimeras in the preservation of electronic records.' *D-Lib Magazine*, 5 (4), April 1999. http://www.dlib.org/dlib/april99/bearman/04bearman.html

Seamus Ross, *Changing Trains at Wigan: Digital Preservation and the Future of Scholarship* London: National Preservation Office, 2000,
http://eprints.erpanet.org/45/01/seamusross_wigan_paper.pdf

Mary Feeney, ed., *Digital culture: maximising the nation's investment*. London: National Preservation Office, 1999. http://www.ukoln.ac.uk/services/elib/papers/other/jisc-npo-dig/

Metadata

Michael Day, *Metadata for preservation*. Cedars project document AIW01. Bath: UKOLN, UK Office for Library and Information Networking, 1998. http://www.ukoln.ac.uk/metadata/cedars/AIW01.html

Disaster management

Seamus Ross and Ann Gow, *Digital archaeology: rescuing neglected and damaged data resources*. London: South Bank University, Library Information Technology Centre, February 1999. http://www.hatii.arts.gla.ac.uk/Projects/BrLibrary/

Portals

Portals exist in both a general and in a technical sense. For the general user, the term portal can refer to any Internet site that acts as a consolidated gateway to desired sources of information. From a technical point of view, a portal is a specific application of certain development tools, allowing a number of different databases and/or applications to be accessed from a single web-based client.

Some examples:

Intute, while only a portal in the general sense of the term, provides access to the best arts and humanities Web resources for education and research, selected and evaluated by a network of subject specialists.

The National Library of Congress Global Gateways project links library resources across the globe (in co-operation with international partners) on specific topics.

Appendix B: AHRC technical appendix

The following explanatory notes are taken from part of the technical appendix that must accompany every application for research funding to the UK Arts and Humanities Research Council (AHRC). IMC would like to thank the AHRC for allowing us to reproduce this document here.

IMC suggests the inclusion of such an appendix in all relevant applications for funding from equivalent Irish research funding bodies. Copies of successful applications could be deposited with a national trusted digital repository, such as the Irish National Data Centre proposed in this document.

Arts & Humanities Research Council

Technical Appendix to research applications

Please complete this appendix if a significant product or by-product of your project is the creation of an electronic resource(s). In advance of completing this appendix, you should read the **'details of the research grant scheme'**, '**guidance notes for the completing and submitting the RG application form'** and AHDS guidelines for AHRC applicants (http://www.ahds.ac.uk/ahrc).

7. Technical appendix guidance notes

If a significant product or by-product of your project is the creation of an electronic resource, you must complete the technical appendix to the **'Research Grants application form'**. This will be assessed by technical reviewers for the Council.

You should ensure that you provide sufficient information on the technical aspects of the proposal to allow the Technical Appendix to be assessed.

The technical appendix has been designed to enable the AHRC to review the technical feasibility of the project. The results of the review will be made available to the peer reviewers who determine the final grade for your application.

The Council is eager to encourage such proposals to adopt community-agreed standards and best practice and to make adequate provision for the long-term preservation, access and integration of the digital resources they create. The AHRC and the Arts and Humanities Data Service (AHDS) have formed a strategic partnership to promote their shared aims with regard to the application of Information and Communication Technology (ICT) in the arts and humanities.

You are strongly advised to consult the guidelines from the AHDS (http://www.ahds.ac.uk/ahrc), and to speak to the relevant AHDS Centre, before submitting an application. For further information please refer to the **'Details of Research Grants scheme'.**

The technical appendix is divided into six sections. Issues which you should consider and address in each section of the appendix are:

1. ***Project management of technical aspects****:* you should indicate how you will manage the technical aspects of the project to ensure its timely and successful completion, and clearly state the electronic output(s). Please note that the scheme of research should outline how the project as a whole will be managed. In particular you should address the following:

    a. management and reporting structure

    b. project timetable

    c. project deliverables

    d. monitoring process.

In addition, you may also wish to describe how the technical aspects of the project have been/ will be piloted including the monitoring and evaluation of the pilot.

2. ***Data development methods****:* in describing data resource development methods you should demonstrate your knowledge and application of best practice. You may wish to focus on some or all of the following issues:

    a. content selection

    b. data/file formats

    c. documenting the resource

    d. advice sought on planning your proposed project

    e. consultation with projects using similar methods

3. ***Infrastructural support****:*

    a. describe the hardware, software and relevant technical expertise that is available to you

    b. indicate what additional hardware, software and relevant technical expertise, support and training is likely to be needed and how it will be acquired

    c. describe the backup procedures that your project will use to safeguard your electronic resource during its development.

4. ***Data preservation and sustainability****:*

    a. you should demonstrate that you have sought advice on any issues which apply to the resource and its preservation.

    b. You should indicate what plans you have to preserve the data, either with the AHDS or through some alternative mechanism.

c. You should also demonstrate the sustainability of the electronic resource created by the project.

5. ***Access****:* you should demonstrate that you have sought advice on and addressed all issues of access. You should indicate what plans you have to make the data/resource available, either with the AHDS or through some alternative mechanism.

6. ***Copyright and intellectual property issues****:* you should demonstrate that you have sought advice on and addressed all copyright and rights management issues which apply to the resource.

In addition you should note requirements for deposit outlined in the AHRC Annexe to the Terms and Conditions of Research Council Grants.

Appendix C: Seminar programme

Digitisation: developing a best-practice template

IAA Lecture Theatre, 28 April 2007

Programme

9.30–10.00: **REGISTRATION, COFFEE & TEA**

Welcome by the Chairperson of the Irish Manuscripts Commission, Mr James McGuire

Session I, Chairperson: Catriona Crowe (National Archives of Ireland)

10.00–10.30 **Best practice models: USA**
Susan Schriebman
Head of Digital Collections and Research, University of Maryland Libraries.

10.40–11.10 **Best practice models: Canada**
Geneviève Allard
Director of Web Content and Services Division, Library and Archives Canada.

11.20–11.40: COFFEE & TEA

Session II, Chairperson: Maurice Bric (University College, Dublin)

11.40–12.10 **Best practice models: Europe**
Dirk Roorda
Project Leader, Data Archiving and Network Services, Netherlands.

12.20–12.50 **Partners in digitisation: Commercial access models**
Bill Kipp, *Consumer Operations Manager (EMEA), Google, Ireland*

13.00–13.30: LUNCH

Session III, Chairperson: Thomas O'Connor (NUI Maynooth)

13.30–14.00 **Long-term digital preservation**
Seamus Ross
Director, Humanities Advanced Technology and Information Institute, University of Glasgow, Scotland.

14.30 **CONCLUDING DISCUSSION & REMARKS**
Jane Ohlmeyer, *Chairperson of the IMC Digitisation Task Force*

Appendix D: Glossary

The following is a glossary of terms used in this report and an explanation of several common acronyms.

Acronym	Expansion (and further details)
ACLS	**American Council of Learned Societies** The ACLS is a private non-profit federation of sixty-eight scholarly organisations in the US. The mission of the ACLS is to advance humanistic studies in all fields of learning in the humanities and the social sciences and to maintain and strengthen relations among the national societies devoted to such studies. They recently published a milestone report on cyberinfrastructure for the humanities and social sciences.
AHDS	**Arts and Humanities Data Service** The AHDS is a UK national data service supported by JISC and AHRC to collect, preserve and promote the electronic resources that result from research and teaching in the arts and humanities.
AHRC	**Arts and Humanities Research Council** UK funding body for third-level research.
Born digital	Digital materials that are not intended to have an analogue equivalent, either as the originating source or as a result of conversion to analogue form. This term is used to differentiate them from (1) digital materials which have been created as a result of converting analogue originals; and (2) digital materials, which may have originated from a digital source but have been printed to paper, e.g. some electronic records.
CDWA	**Categories for the Description of Works of Art** Devised at the J. Paul Getty Trust, CDWA is a metadata set for the description of works of art. It is a data dictionary only (that is, it provides a list of data elements which could be relevant to the description of artworks, and corresponding definitions). It is quite a detailed and extensive list, but it applies only to artworks. There is no structure, no syntax, and no standards on what the content should be or how it should be formed.
CESSDA	**Council of European Social Science Data Archives** CESSDA is an umbrella organisation for social science data archives across Europe. It promotes the acquisition, archiving and distribution of electronic data for social science teaching and research in Europe.
CNRS	**Centre National de la Recherche Scientifique** **CNRS** is the French data centre service partner in DARIAH project of the ESFRI.
CURL	**Consortium of University Research Libraries** **CURL**'s mission is to increase the ability of research libraries to share resources for the benefit of the local, national and international research community. CURL and JISC jointly published a report on the need for a

Acronym	Expansion (and further details)
	national framework for digitisation in the UK.
DANS	**Data Archiving and Networked Services** Comparable with the AHDS, **DANS** is the Dutch national organisation responsible for storing and providing permanent access to research data from the humanities and social sciences.
DAREnet	DAREnet is the Dutch network of Digital Academic REpositories. It is a result of the national DARE programme involving all Dutch universities and several academic and research organisations. Its aim is to coordinate and stimulate a network of Dutch repositories containing information from academic research. It thus enhances the accessibility to and visibility of this information, both national and international. DARE is a national initiative coordinated by the SURF Foundation. All DARE participants have their own repository, for which they are individually responsible. These repositories are brought together under the DARE banner and collectively form the building blocks of the DARE network.
DARIAH	**Digital Research Infrastructure for the Arts and Humanities** A project of the Social Sciences and Humanities division of the ESFRI that proposes to provide a European working structure for the co-ordination of digitisation infrastructure and projects. It will be based on an existing network of data centres in the UK, the Netherlands, Germany and France.
Dark Archive	An archive with no physical connection to external computer networks. This represents the most secure form of long term preservation for archived digital objects. An exemplary model is represented by Florida State's Dark Archive (http://www.fcla.edu/digitalArchive/pubs.htm), recently audited and pronounced excellent.
DAST	**Department of Arts, Sport and Tourism** The Irish government department with responsibility for the national cultural institutions and the Irish Manuscripts Commission.
DHO	**Digital Humanities Observatory** An RIA proposal for a repository for the arts and humanities feeding into a digital infrastructure for Ireland that will interface with similar European data centres and open up Irish research data (in the first instance) to a wider community.
DRIVER	Digital Repository Infrastructure Vision for European Research DRIVER sets out to build the test bed for a future knowledge infrastructure of the European Research Area. Aimed to be complimentary to GEANT2, the successful infrastructure for computing resources, data storage and data transport, DRIVER will deliver the content resources, i.e. any form of scientific output, including scientific/technical reports, working papers, pre-prints, articles and original research data. The vision, to be accomplished in a second phase, is to establish the successful interoperation of both data network and knowledge repositories as integral parts of the E-infrastructure for research and education in Europe.
EROHS	**European Resource Observatory for the Humanities and the Social Sciences** A project proposal for the 'roadmap' for the Humanities and Social

Acronym	Expansion (and further details)
	Sciences division of the ESFRI.
ESFRI	**European Strategy Forum on Research Infrastructures** ESFRI is an EU vehicle for consultation on strategic issues relating to the long-term development of research infrastructures that will support achieving the knowledge growth as set out in the Lisbon strategy. Since 2004 it has worked on producing a 'roadmap' document for the construction of the next generation of large-scale research infrastructures across seven discipline-groups from social sciences and the humanities to astronomy, astrophysics, nuclear and particulate physics. Within each group, core projects that will move the ideals of the 'roadmap' forward have been identified.
Folksonomy	A folksonomy is a user-generated taxonomy used to categorise and retrieve web content such as Web pages, photographs and Web links, using open-ended labels called tags. Typically, folksonomies are Internet-based, but their use may occur in other contexts. The folksonomic tagging is intended to make a body of information increasingly easy to search, discover, and navigate over time. A well-developed folksonomy is ideally accessible as a shared vocabulary that is both originated by, and familiar to, its primary users. Two widely cited examples of websites using folksonomic tagging are Flickr and del.icio.us
HII	**Humanities Institute of Ireland** Based at University College, Dublin, HII was established in 2002 under the Higher Education Authority's Programme for Research in Third Level Institutions (PRTLI). It is a unique locus for leading-edge interdisciplinary research in the humanities and social sciences not only within the university itself but also nationally. The HII is currently funded to undertake a major team-based research programme on issues on identity, memory and meaning in the twenty-first century.
i2010 and i2010 Digital Library Project	**i2010** is the European Commission's strategic policy framework laying out broad policy guidelines for the information society and the media in the years up to 2010. It promotes an open and competitive digital economy, research into information and communication technologies, as well as their application to improve social inclusion, public services and quality of life. The **Digital Libraries Initiative** is a flagship project of the Commission's i2010 strategy. The initiative aims at making Europe's diverse cultural and scientific heritage (books, films, maps, photographs, music, etc.) easier and more interesting to use online for work, leisure and/or study. It builds on Europe's rich heritage combining multicultural and multilingual environments with technological advances and new business models.
ICT	Information and communication technology
Institutional Repository	A collection point for digital objects (theses, published research papers, images, sound archives, library and archive resources and teaching resources), for example, TARA (Trinity's Access to Research Archive), DCU Institutional Repository, NUI Maynooth ePrints and eTheses Archive and the National Digital Learning Repository and the Irish Virtual Research Library and Archive (IVRLA). **Equivalent to the Local Repository in the schematic diagram on page 25.**

Acronym	Expansion (and further details)
ISSDA	**Irish Social Science Data Archive** ISSDA holds and makes available data from opinion surveys and official statistics (such as the Census). It operates within the Geary Institute, UCD and is managed jointly by UCD and Economic and Social Research Institute, with the collaboration and support of the Central Statistics Office. It was established following the publication of The Irish Data Archive feasibility Project: A report by the Data Archive, University of Essex, by the Social Research Council of the RIA.
IVRLA	**Irish Virtual Research Library & Archive** IVRLA is a major digitisation and digital object management project launched in UCD in January 2005 and funded under PRTLI as part of the HII. Its primary remit is to preserve elements of UCD's main repositories and increase and facilitate access to this material through the adoption of digitisation technologies. The project will also undertake dedicated research into the area of interacting with, and enhancing the use of, digital objects in a research environment through the development of a digital repository. When fully implemented, the IVRLA will be one of the first comprehensive digital primary source repositories in Ireland.
JISC	**Joint Information Systems Committee** JISC's activities support education and research in the UK by promoting innovation in new technologies and by the central support of Information and Communication Technology services.
Lund Principles	On 4 April 2001, under the Swedish EU-Presidency, the European Commission organised an expert meeting with representatives from all Member States in Lund. The conclusions and recommendations derived from this meeting are known as the Lund Principles and were further developed in the Lund Action Plan, which establishes an agenda for actions to be carried out by Member States and the Commission. The main conclusions at Lund were for Member States to: establish an evolving forum of coordination; support the development of a European view on digitisation policies and programmes; develop mechanisms to promote good practice and skills development and; collaborate to make the digitised cultural and scientific heritage of Europe visible and accessible. This is being realised through the National Representatives Group (NRG). The successor of the Lund Action Plan was presented under the UK Presidency: Dynamic Action Plan for the EU coordination of digitisation of cultural and scientific content
Metadata	Metadata facilitates the finding of information by the assigning of consistent descriptors to information resources. It is important to conform to European and international guidelines for best practice when designing metadata fields in a project. This will ensure interoperability between projects nationally and internationally. See Appendix A: Further reading for a list of metadata standards.
MINERVA(EC)	**MInisterial NEtwoRk for Valorising Activities in digitisation, eContentplus - Supporting the European Digital Library**. MinervaEC is a Thematic Network in the area of cultural, scientific information and scholarly content. It brings together stakeholders and experts from all over Europe, capitalising on the results achieved by the

Acronym	Expansion (and further details)
	previous Minerva project, and supporting the European Commission initiative “i2010 – A European Information Society for growth and employment” as well as the Dynamic Action Plan launched in Bristol in November 2005 by the European Union Member States. MinervaEC will operate through the co-ordination of national policies, programmes and institutions of the cultural sector, and by supporting the National Representatives Group of the European Ministries of Culture. Its goal is to facilitate the creation of added value products and services at European level, to improve awareness of the state-of-the-art in the sector, to overcome the fragmentation and duplication of digitisation activities of cultural and scientific content and to maximise co-operation among the Member States.
Mirror site	A Web site that is a replica of an already existing site, used to reduce network traffic (hits on a server) or improve the availability of the original site. Mirror sites are useful when the original site generates too much traffic for a single server to support. Mirror sites also increase the speed with which files or Web sites can be accessed: users can download files more quickly from a server that is geographically closer to them.
NINCH	**National Initiative for a Networked Cultural Heritage** NINCH is a diverse nonprofit coalition of arts, humanities and social science organisations created to assure leadership from the cultural community in the evolution of the digital environment. The Initiative began in 1993 as a collaborative project of the American Council of Learned Societies, the Coalition for Networked Information, and the Getty Information Institute, an operating program of the J. Paul Getty Trust.
NINCH Guide	The NINCH **Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials** is a practical online guide for those in all sectors of the community who are digitising and networking cultural resources. A NINCH Working Group representing all sectors of the community created a set of six core principles defining good practice. The group then outlined the scope of a Guide that would be based on a survey of current practice and organized as a decision tree for the user.
NDLR	**National Digital Learning Repository** The NDLR is a HEA-funded pilot project between Irish Universities and the Institutes of Technology to support the collaboration and sharing of learning and teaching resources. The NDLR project is investigating a framework to enable development and sharing of digital learning resources between the seven universities of Ireland, Dublin Institute of Technology and the thirteen institutes of technology.
NRG	**National Representatives Group** National Representatives Group (NRG), made up of officially nominated experts from each Member State, has been established to act as "a steering group for the activities related to the co-ordination of digitisation policies and programmes, with special emphasis on cultural and scientific resources and on the contribution of public cultural institutions." Its stated mission is to monitor progress regarding the objectives encapsulated in the Lund Principles.
OAI-PMH	**Open Access Initiative Protocol for Metadata Harvesting** OAI-PMH is a framework for application-independent interoperability based

Acronym	Expansion (and further details)
	on metadata harvesting. Widely accepted internationally, it ensures interoperability of repositories and exposure of metadata to web harvesters, crawlers and search engines such as Google.
OAIS	**Open Archives Information System** As an ISO standard OAIS advocates and promotes a particular functional and informational model, which will support the long-term preservation and presentation of digital and electronic sources.
Portal	Term used to describe a website providing a mechanism for opening up digitised data on a common subject from a variety of sources. The collections/data/catalogues appear to the user to be from the one source.
Preservation	In terms of digitisation, preservation is the maintenance of digital data in a readable format and the migration of data to new formats when old formats/methods of reading them become obsolete.
PRTLI	Programme for Research in Third Level Institutions, a programme of the Higher Education Authority (HEA).
RIA	**Royal Irish Academy** The RIA is an all-Ireland, independent academic body that promotes study and excellence in the sciences, humanities and social sciences.
RLG	**Research Libraries Group** A not-for-profit membership corporation of institutions devoted to improving access to information that supports research and learning.
SIF	Strategic Innovation Fund, a programme for funding in the Higher Education Authority (HEA).
SURF	SURF is the collaborative organisation for higher education institutions and research institutes in the Netherlands. It is aimed at breakthrough innovations in ICT and provides the foundation for excellence in higher education and research. SURF comprises of SURFfoundation, SURFnet and SURFdiensten (SURFservices)
TDR	**Trusted Digital Repository** As defined by RLG in 2002, a trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future. The RLG has produced a framework of attributes and responsibilities for trusted, reliable, sustainable digital repositories capable of handling the range of materials held by large and small research institutions. The framework is broad enough to accommodate different situations, architectures, and institutional responsibilities while providing a basis for the expectations of a trusted repository. The critical component will be the ability to prove reliability and trustworthiness over time.
UKOLN	**UK Office for Library Networking** UKOLN is a centre of expertise in digital information management, providing advice and services (including standards) to the library, information, education and cultural heritage communities.

Acronym	Expansion (and further details)
VRA Core	VRA Core v 4.0 is a metadata standard for the cultural heritage community that was developed by the Visual Resources Association's Data Standards Committee.

Appendix E: Repository software options

The following has been provided by John McDonough (Irish Virtual Research Library and Archive, UCD) and was drawn from documentation created by the IMPROVE Fedora consortium and Sandy Payette of Cornell University. IMC records its appreciation for this additional material for its policy document.

Repository architecture

Repositories are a core building block of the scientific knowledge domain. They allow scholars and scientists to deposit material, work on it, interrelate it, thus creating new scholarship and knowledge. By doing so, they create valuable assets, which need to be preserved for the long term. Scientists, scholars, content managers, and librarians need a flexible content repository system that allows them to uniformly store, manage, and deliver all their existing content that also accommodates new forms that will inevitably arise in the future.

Besides many organisational, political and legal aspects, the capabilities of the software—and its architecture—used to develop digital repositories are critical. They determine how many of the important characteristics of the repository will be implemented, for example, interoperability of the repository, the persistency of data and the relations between data objects.

Unfortunately, the digital libraries and institutional repositories built up in recent years are often based on proprietary software and closed standards. These types of independent, vertically integrated library systems do not meet the requirements of scientific communication in the age of worldwide e-Science. Further, the value added principle of networked collaboration remains unused if institutional repositories and digital libraries continue to only use workflows for producing customised research objects. Therefore, only open and standards-based repositories are to be considered for the European e-Infrastructures.

Open source software for building repositories

Today, several open source software packages for managing digital repositories exist. The OSI Guide to Institutional Repository Software mentions the most important ones: ARNO, CDS Invenio, DSpace, EPrints, Fedora, and i-Tor. A study from New Zealand evaluated all six systems, but found only three of them worthwhile an in-depth analysis: DSpace, EPrints, and Fedora. Of the remaining systems, CDS Invenio and i-Tor have very little community surrounding them and only few known installations, which raises concerns about the sustainability of the software. ARNO relies on a non-open source database management system.

DSpace

MIT's DSpace was expressly created as a digital repository to capture the intellectual output of multidisciplinary research organisations. It integrates a user community orientation into the system's structure. This design supports the participation of the schools, departments, research centres, and other units typical of a large research institution. DSpace is also focused on the problem of long-term preservation of deposited research material. The system is easy to setup and deploy, provides an out-of-the-box experience, and has a very open development community.

EPrints

EPrints is a good candidate for many institutions as it is the least complex of the three systems, and hence has the lowest skill level barrier of the three to implement and maintain. EPrints has the widest install base, a significant factor in that it goes a long way to ensure its longevity as a fully supported system.

The data model causes some scalability issues. Its method of adding new digital content type can lead to disparate data models and compatibility issues if maintaining multiple systems. The development of the software is done in a closed community at the University of Southampton, which retains the copyright to EPrints.

Fedora

Fedora demonstrates the best scalability among the three short-listed systems. It utilises a very flexible digital object model and can store multiple types of digital objects and collections particularly well. It has a strong development team and development roadmap. As a foundation-architecture with powerful API based interoperability features, Fedora is highly flexible and powerful. Unlike the other systems Fedora is not constrained to fulfil a particular function and a variety of front ends (user-interfaces together with tools) can be integrated or developed for the utilisation of the architecture depending upon particular requirements of an institution or community.