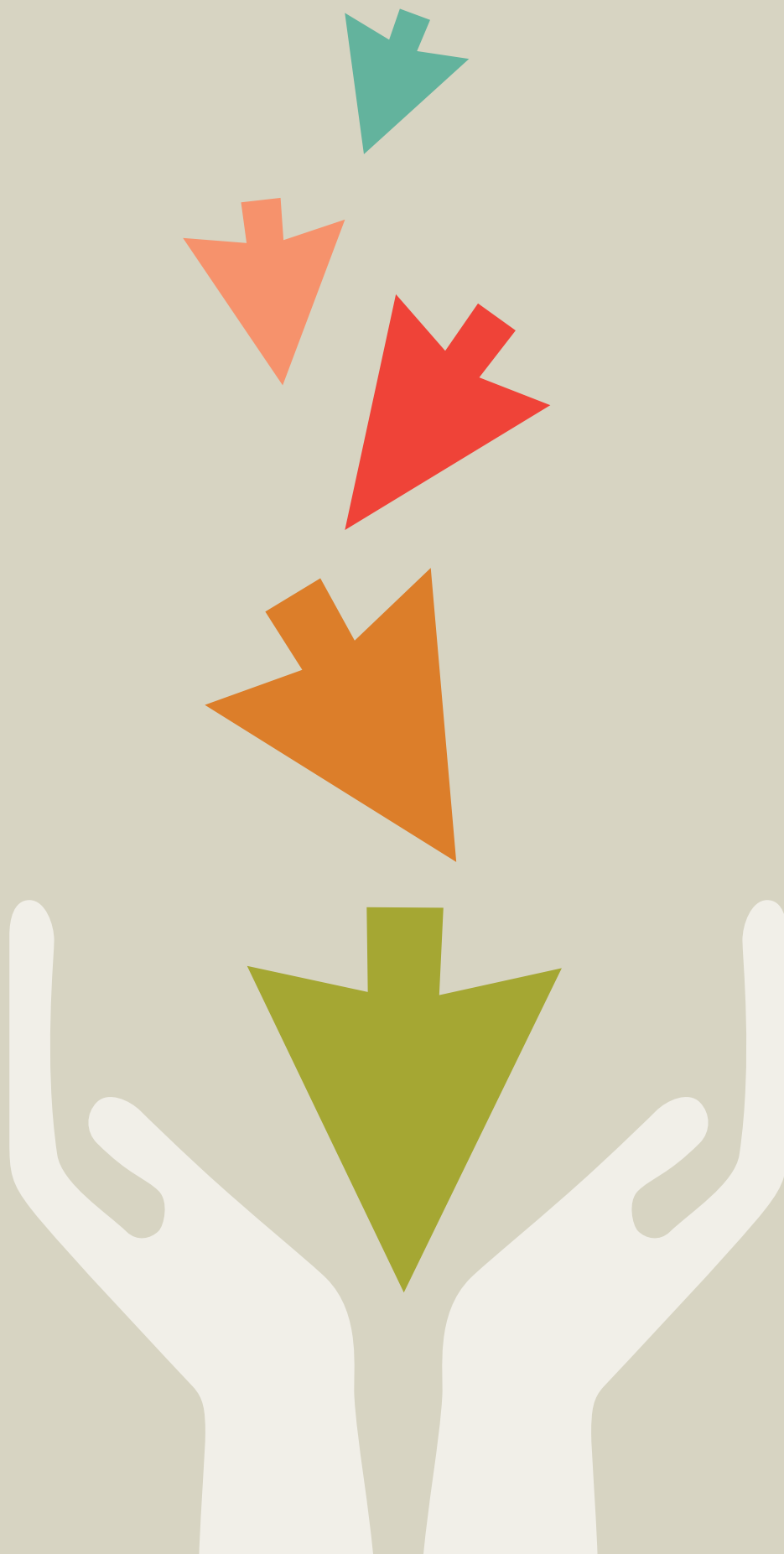


Caring for Digital Content

Mapping International Approaches



Caring for Digital Content

Mapping International Approaches



First published in 2013 by the Royal Irish Academy

© NUI Maynooth, Trinity College Dublin, and the Royal Irish Academy

When citing this report please use the following citation: O'Carroll, A., Collins, S., Gallagher, D., Tang, J., & Webb, S. (2013) *Caring for Digital Content, Mapping International Approaches*. Maynooth: NUI Maynooth; Dublin: Trinity College Dublin; Dublin: Royal Irish Academy. DOI: 10.3318/DRI.2013.1

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any electronic, mechanical or any other means, now known or hereafter invented, including photocopying and recording, or otherwise without either the prior written consent of the publishers or a licence permitting restricted copying in Ireland issued by the Irish Copyright Licensing Agency Ltd, The Writers' Centre, 19 Parnell Square, Dublin 1.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

Digital Repository of Ireland series, no. 2

Print ISSN 2009-6445

Online ISSN 2009-6461

ISBN 978-1-908996-25-1

Design: Fidelma Slattery

Contents

6	Director's Foreword
7	The Digital Repository of Ireland
8	Introduction
10	Metadata Aggregators
12	Single-Site Digital Repositories
15	Multi-Site Digital Repositories
17	Information Providers
19	Membership Organisations
22	EU Infrastructure Projects
23	Digital Projects: Ongoing
25	Digital Projects: Completed
27	Conclusion

Director's Foreword



DR SANDRA COLLINS

The Digital Repository of Ireland (DRI) has carried out a mapping exercise to identify exemplars of international repositories, repository projects and organisations with expertise in the management of digital data.

It is important for DRI to have an understanding of international and national digital preservation and access practices in order to inform our requirements specification in building the national trusted digital repository.

We believe it is critical to be informed by, and to inform, international best practice. We should learn from each other's experiences. In this manner we can work with the community to advance the state of the art for the collective good, and it is in this spirit that we publish our findings to date.

This report is the result of our international mapping exercise, and it complements our national report *Digital Archiving in Ireland: National Survey of the Humanities and Social Sciences*,¹ co-authored by Aileen O'Carroll and Sharon Webb. It also forms part of our longer-term goal to publish national guidelines for digital preservation and access that will be designed with the community for use by the community.

The mapping is not exhaustive and represents a snapshot in time. We shall continue to monitor international developments and identify emerging practices and organisations, and in turn contribute to the national and international landscape.

Digital preservation of our social and cultural heritage is imperative, and this is a step towards that goal.

¹ The report is available at http://dri.ie/sites/default/files/files/Digital_Archiving_In_Ireland_2012.pdf (accessed 12 January 2013).

The Digital Repository of Ireland

The Digital Repository of Ireland (DRI) is an interactive, trusted digital repository for social and cultural content held by Irish institutions. By providing a central internet access point and interactive multimedia tools, the DRI facilitates engagement with contemporary and historical data, allowing the public, students, and scholars to research Ireland's cultural heritage and social life in ways never before possible. As a national digital infrastructure, the DRI is working with a wide range of institutional stakeholders to link together and preserve Ireland's rich and varied humanities and social science data.

The DRI is also acting as a focal point for digital best practices by collaborating on the development of guidelines, and working to inform national policy on digital preservation and access. The DRI promotes awareness of the benefits of preservation and open access to data, while respecting the importance of ownership, copyright, intellectual property, privacy and confidentiality. DRI has developed a lean repository prototype, organised a major international 280-participant symposium, and published a national report with the findings from our nationwide programme of stakeholder interviews to determine the digital preservation and access practices in cultural institutions, libraries, higher education institutions and funding agencies. DRI seeks to share best practices with the community to enable cost savings and improved standards of preservation and access.

The DRI was launched in 2011, when it received funding from the Irish Government's PRTLI cycle 5 for €5.2M over four years. The DRI consortium is composed of the following partners: The Royal Irish Academy (Lead Institution), National University of Ireland Maynooth (NUIM), Trinity College Dublin (TCD), Dublin Institute of Technology (DIT), National University of Ireland Galway (NUIG), and National College of Art and Design (NCAD). The DRI Research Consortium are currently collaborating with a network of cultural, social, academic and industry partners, including the National Library of Ireland (NLI) and the Irish National Broadcaster RTÉ.

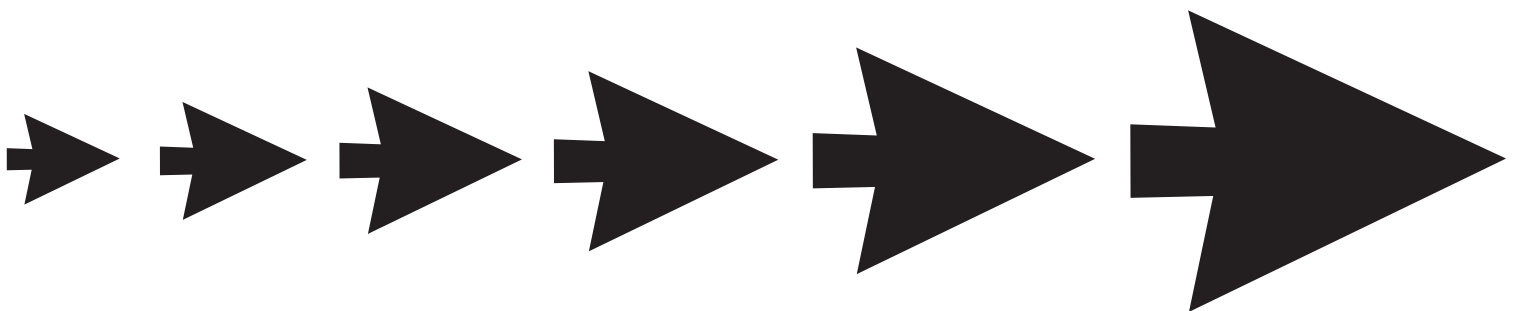


² The report is available at http://dri.ie/sites/default/files/files/Digital_Archiving_In_Ireland_2012.pdf (accessed 12 January 2013).

Introduction

Digital data is increasing. In 2007 it was estimated that the digital world was 291 gigabytes. At that point the amount of information ‘created, captured or replicated exceeded the available storage’.³ Increasingly humanities and social science data is created in digital form. This document maps the emerging approaches to caring for digital data. The repositories included in this mapping were not selected on the basis of their content (for example, whether they cared for datasets rather than research outputs) or their research field (that is, whether they emerged from libraries, cultural institutions or social scientific bodies). Rather they were selected on the basis of the approach they took to caring for digital content. Following a review of international institutions, three different models emerged: the metadata aggregator, the single-site repository and the multi-site repository.

This report documents examples from each model. The list, however, is not exhaustive. Repositories were selected if the project was currently ongoing, with funding that was either large-scale or expected to persist in the future. The inclusion of a number of partner institutions (particularly Irish partners) was also a factor that led to selection for this report. Representatives of European and US repositories were included. However, the selection is biased towards repositories for which English-language information was available. This report includes many of the key developments in the field, and as such serves as a guide to those who are seeking information (particularly in English) on the range of approaches to caring for digital data.



Additionally the report documents key organisations and information providers that collate domain-specific expertise in the field. The seeds of future developments in the field of digital preservation are often to be found in EU projects. For this reason this mapping document concludes with a brief overview of past and current EU digital preservation projects.

³ The quote is from Gantz, J. et al. (2008) *The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth through 2011*. IDC White Paper. Framingham, MA: IDC

The DRI is envisaged as a federated infrastructure at the technical level; there are a number of levels that can be federated to preserve and manage the data that is stored. The definition of 'federation' is an organisation or group within which smaller divisions have some degree of internal autonomy. Federation can occur at a number of layers in the software and hardware infrastructure: not only at a technological level but also at the organisational levels. Different organisations will have different models of federation, distribution and multi-site definitions. This mapping of international approaches has been undertaken in order to inform the decisions that DRI will make with respect to a federated infrastructure.

The DRI is envisaged as a federated infrastructure at the technical level; there are a number of levels that can be federated to preserve and manage the data that is stored.



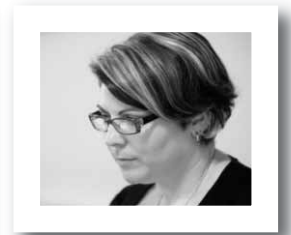
Dr Aileen O'Carroll (NUIM)



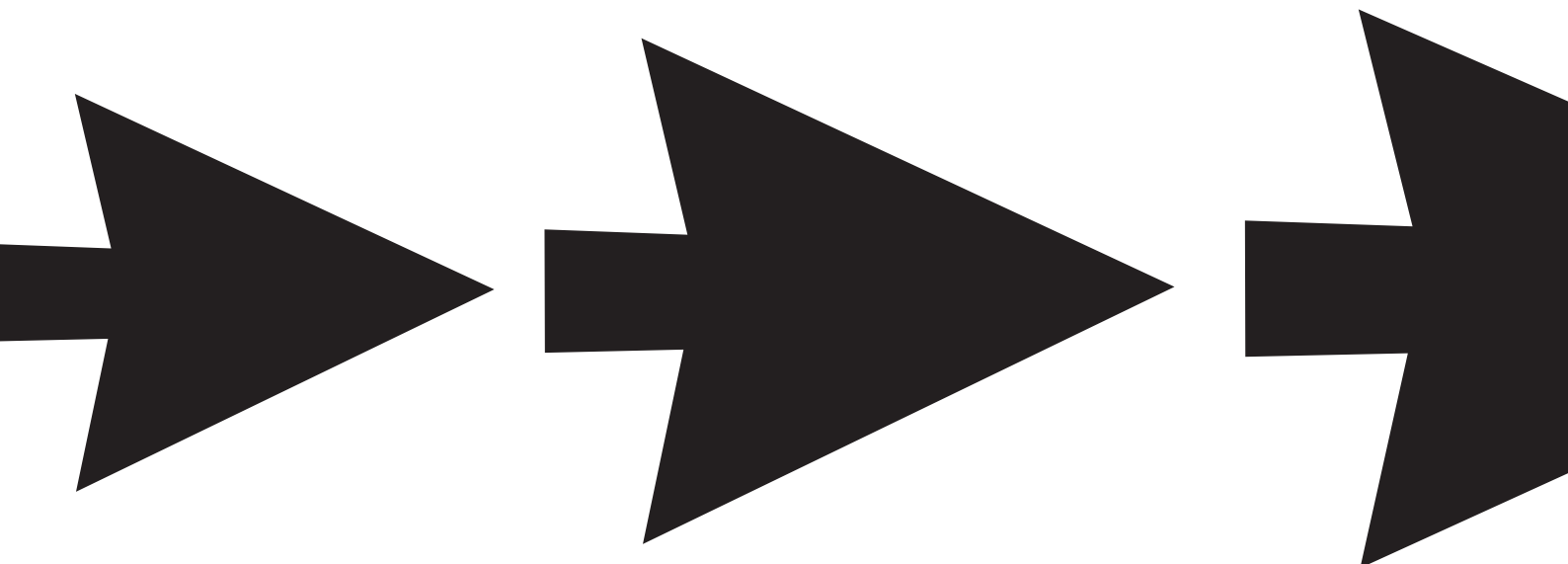
Damien Gallagher (NUIM)



Jimmy Tang (TCD)



Dr Sharon Webb (NUIM)



Metadata Aggregators

One approach to making digital data more widely available is to aggregate the metadata from existing repositories. This serves to highlight local collections and build thematic links between data located in different institutions. Here we discuss the following examples:

- **Australian National Data Service** – <http://ands.org.au>
- **Digital Public Library of America** – <http://dp.la>
- **Europeana** – <http://www.europeana.eu/portal>
- **EUScreen** – <http://www.euscreen.eu/index.html>
- **Hathi Trust** – <http://www.hathitrust.org>

Australian National Data Service

The Australian National Data Service was established in 2007. It consists of a national network of data repositories with a single access interface that allows the user to browse Australian research datasets and projects. Its discovery service gives access to information about 41,520 datasets and over 30,000 projects or programs that create datasets.

Digital Public Library of America

The DPLA offers a single point of access to 2.5 million metadata records from US based libraries, archives and museums. It is funded by government and foundation grants, including the Alfred P. Sloan Foundation. It is establishing a US network of over 40 state and regional digital libraries, as well as large digital libraries, museums, archives and repositories. They have more than 450 partners.

Europeana

Europeana is primarily an aggregator of metadata using the ESE standard. There are currently over 150 partners in the project, and a number of Irish institutions contribute metadata to Europeana. Libraries, museums and multimedia libraries aggregate metadata into one website. It has developed the Europeana portal through which users can access content from various digital repositories. It has also developed search and browsing interfaces and metadata schemas for a number of specific projects that bring

together similar data-types. It was launched in 2008 and plans to run until 2025. At the moment it facilitates online access to some 20 million digitised objects, including books, maps, newspapers, journals, photographs, sound and video.

Content is hosted by the contributing institutions, not by Europeana itself. Only contextual information is collected, stored and indexed. Contributing sites must organise their data in a specific way such that it can be addressed and searched, however, contributors keep full control over their content. The project is currently aiming to exploit, reuse and develop open source technologies to deliver Europeana 1.0.

The Europeana project faces a number of challenges that involve collecting, indexing and doing quality assurance that the data meets the prescribed Europeana standards. The consortium provides a range of XSLT files for translating various metadata formats to their required harvesting format.

Similarly to Europeana, DRI intends to aggregate and harvest metadata that relate to its own mission objective.

EUscreen

EUscreen is a multilingual portal that allows online access to videos, stills, texts and audio from European broadcasters and audiovisual archives dating from the early 1900s to the present day. The portal was launched in 2011 and connects with Europeana. A consortium of 36 partners and associate partners from 19 countries, including the RTE Archives, has developed the project. It received funding from the EU eContentplus programme from 2009 for three years. It is intended to allow access to over 30,000 items of programme content and information.

Hathi Trust Digital Library

Hathi Trust, initially a US-based partnership of 13 universities, was founded in 2008 with the intent of setting up a repository to archive and share their digital collections. It received funding of 1.3 million dollars from the Andrew W. Mellon Foundation. The initial aspiration of the partnership has been to preserve and provide access to the digitised content from the partner library collections, Hathi Trust initiatives, as well as public domain content digitised by Google, the Internet Archive, and Microsoft. By July 2013, 10,733,126 volumes had been added to The Hathi Trust digital library.



Single-Site Digital Repositories

Single site repositories host databases and the associated functions of archiving, including data preparation and preservation, within one location. Many national repositories are single-site repositories. Many will locate off-site back-up in multiple locations, but the main technological infrastructure is located in one site. Examples include:

- **National Library of New Zealand** – <http://www.natlib.govt.nz>
- **Data Archiving and Network Services (DANS)**, Netherlands – <http://www.dans.knaw.nl/en>
- **The Internet Archive**, California – <http://www.archive.org>
- **Multimedia Educational Resource for Learning and Online Teaching (MERLOT)**, California – <http://fedsearch.merlot.org/fedsearch/fedsearch.jsp>
- **The Netherlands Institute for Sound and Vision** – <http://www.beeldengeluid.nl/en>
- **UK Data Archive (UKDA)** – <http://www.data-archive.ac.uk>

National Library of New Zealand

The role and purpose of the National Library of New Zealand is defined by the National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003. Its role is to collect, preserve and protect artefacts that relate to New Zealand and to disseminate this information to the people of New Zealand. It develops and maintains a number of collections, provides access and promotes co-operation and collaboration with others both in New Zealand and overseas. New Zealand has a 10-year strategic plan for digitisation. Digital New Zealand was launched in 2008.

Data Archiving and Network Services (DANS)

Established in 2005, DANS is the Dutch national archive of digital research data. Its mission is to promote sustained access to digital research data. It provides archiving services, training in reuse of data and outreach. It is also the home of the Data Seal of Approval, an accreditation guideline for digital repositories.

The Internet Archive

Founded in 1996 and based in San Francisco, the Internet Archive is a non-profit organisation with the aim of building an Internet Library. Its purposes include offering permanent access for researchers, historians, scholars, people with disabilities and the general public to historical collections that exist in digital format. The archive now includes a number of different types of data, such as texts, audio, moving images, software and archived web pages/sites. The primary archive is stored in San Francisco. To ensure the stability and safety of the data, a mirror of the data is maintained at Bibliotheca Alexandrina⁴ in Egypt. The archive allows the public to both upload and download the data from its storage systems, and provides unrestricted access at no cost. The Internet Archive can be seen as an umbrella organisation with a number of preservation and archival projects.

Multimedia Educational Resource for Learning and Online Teaching (MERLOT)

MERLOT is a free and open online community of resources in a number of areas such as business, humanities, social sciences, mathematics and statistics, and arts, based in California State University, Long Beach. The MERLOT system contains user-centred, peer-reviewed and online learning materials that are catalogued by registered members. The project has strategic goals of improving the learning quality and quantity of peer-reviewed material as well as improving the effectiveness of using such material in teaching and learning. At first glance MERLOT appears to be an aggregator of metadata from a number of digital repositories, but it also contains a repository. This repository contains not only collections of learning and teaching material, but also what are referred to as Personal Collections. These appear to be linked collections of data from the overall MERLOT system. The MERLOT Repository allows users to comment on material and Personal Collections that have been stored in the system.

The Netherlands Institute for Sound and Vision

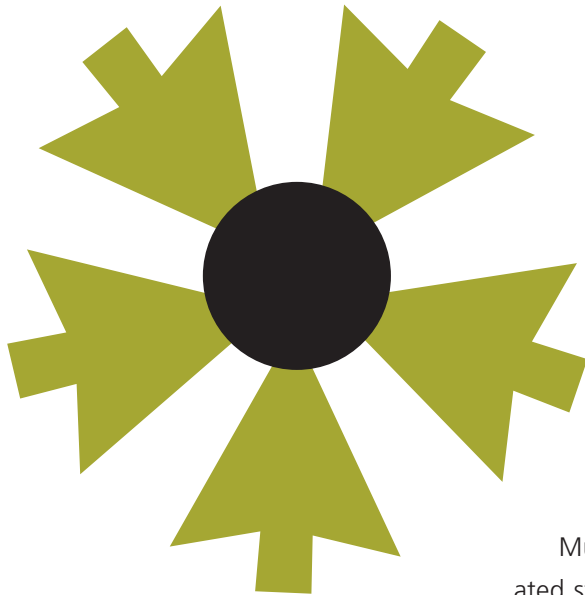
The Netherlands Institute for Sound and Vision is a nationally funded institution that manages over 70 percent of the Dutch audiovisual heritage. The collection contains

⁴ The Bibliotheca Alexandrina website is at <http://www.bibalex.org>

more than 750,000 hours of television, radio, music and film, from 1898 until today. A subsection of the audiovisual collection can be accessed via a web portal. Over the period 2007–2014 it is engaging in a digitisation project at the end of which it is projected that 91,183 hours of video, 22,086 hours of film, 98,734 hours of audio material and over 2.5 million pictures will be digitised and made accessible to the public.

UK Data Archive (UKDA)

UKDA currently curates the UK's largest collection of digital data in the area of social sciences and humanities. It was founded in 1967; the first downloadable datasets became available in 2001. The Economic and Social Research Council (ESRC), JISC (formerly Joint Information Systems Committee) and the University of Essex primarily fund it. UKDA acts as a broker to the collections it is committed to curating, and is a national project in the UK. The project also includes the UK Data Store, an online self-archiving system that collects, curates and preserves a range of digital objects. The project provides documentation and workflows to users on how to access and share data in the system. There are suggestions that the UKDA will move to a more distributed structure in the future.



Multi-Site Digital Repositories

Multi-site repositories host data within a federated structure that allows sharing of metadata and data across institutions. Examples include:

- **ARTstor** – <http://www.artstor.org>
- **Digital Repository of Ireland** – <http://www.dri.ie>
- **IQSS Dataverse Network** – <http://thedata.org>
- **Open Access Infrastructure Research for Europe** (Openaire) – <http://www.openaire.eu>
- **Texas Digital Library** (TDL) – <http://www.tdl.org>

ARTstor

ARTstor was founded by The Andrew W. Mellon Foundation to address the problems faced by institutions of higher education during the 1990s with the migration from physical and analogue to digital media. ARTstor builds on JSTOR,⁵ a similar project that handles archiving of digitised scholarly journals. ARTstor involves a number of partners from China, France, the UK and the US. Its goals are to provide a non-profit service to assemble and build image collections across time periods and countries. ARTstor consists of a repository of images, tools to access and use the images and a restricted usage environment. It provides image management software, known as Shared Self, which enables institutions to manage, store, use, and publish their institutional and faculty image collections within their institution or publicly on the web.

Digital Repository of Ireland

The Digital Repository of Ireland (DRI) is an interactive national trusted digital repository for contemporary and historical, social and cultural data held by Irish institutions; it provides a central internet access point and interactive multimedia tools for use by the

⁵ The JSTOR website is at <http://www.jstor.org>

public, students and scholars. DRI is a four-year Irish exchequer-funded project initiated in 2010, comprising six Irish academic partners, and is supported by the National Library of Ireland, the National Archives of Ireland (NAI) and the Irish national broadcaster RTÉ.

IQSS Dataverse Network

The Institute for Quantitative Social Sciences (IQSS) is based at Harvard University; it is both a research centre and a part of the Harvard administration. The organisation provides services, support and infrastructure to its users. The goals of IQSS are to create, preserve and make widely accessible the collections and tools that are deposited in the system. It hosts the Dataverse Network, an application and service that allows researchers to create their own data archive which is either hosted in the IQSS or located on the user's own technical infrastructure. The Dataverse Network enables sharing of metadata across these data archives and provides citation information on datasets. The source code that is used to build IQSS and its architecture is available publicly.⁶ The IQSS system is available to anyone in the world to contribute to and access. The system's software architecture is built around a tiered framework of Java applications.⁷ If desired, a user can deploy their own version of IQSS's software stack. The IQSS's model of federation is that it delegates the access controls to its users; the systems are primarily located at IQSS's data centres.

Openaire, Openaire Plus

Openaire was Initiated under EU Framework 7 in 2009. Its 41 members from 33 European countries have created a digital repository that facilitates access to the Open Access scientific production of the European Research Area; future developments are intended to provide cross-links from publications to data and funding schemes. Trinity College Dublin was an Openaire partner, participating as the Irish National Open Access liaison office and remains in Openaire Plus which will run until May 2014.

Texas Digital Library

The Texas Digital Library (TDL) is a consortium of higher educational institutes in Texas set up to provide shared services in support of research and teaching. The goal of this organisation is to combine the resources of its members to provide cost-effective solutions; to encourage collaborations; to provide open access to data, preservation of data, a platform for scholars, reduction in cost by sharing expertise and infrastructure; and to create a competitive advantage for seeking funds and grants. TDL provides digital institutional repositories, centrally managed storage, a management and publication system as well as a number of other services related to supporting scholarly research activities.

⁶ Details of the architecture and code are given at <http://thedata.org/book/architecture> and <http://sourceforge.net/projects/dvn/files>

⁷ It utilises Java Server Faces (JSF2) in the User Interface layer, Enterprise Java Beans (EJB3) in the middle tier, PostgreSQL as the database, Lucene as the index server, R and Zelig for the data analysis component.

Information Providers

A number of domain-specific institutions provide advice and support to those engaged in preserving digital data. These include:

- **Databib** – www.databib.org
- **DataONE** – <http://www.dataone.org>
- **Digital Curation Centre** (DCC) – <http://www.dcc.ac.uk>
- **Inter-university Consortium for Political and Social Research** (ICPSR) – <http://www.icpsr.umich.edu/icpsrweb/ICPSR>
- **NESTOR** (Network of Expertise in Long-term Storage of Digital Resources) – <http://tinyurl.com/cnb4kfp>

Databib

Databib is a collaborative, annotated bibliography of primary research data repositories that includes policies for access, use, and deposit of data. It was created by Purdue and Penn State Universities. Users can add further information to the bibliography.

DataONE

DataONE is a US-based research infrastructure, funded by the US National Science Foundation; its aim is to ensure the preservation, access, use and re-use of multi-scale, multi-discipline, and multinational science data.

Digital Curation Centre

The DCC is a UK-based centre that provides advice on data management, digital preservation and archiving best practice. It produces a number of guides and tools, and provides ongoing training courses on aspects of digital content management.

Inter-university Consortium for Political and Social Research

The ICPSR is both a data archive of social science research data and an international consortium of about 700 academic institutions and research organisations. It is based

in the Institute for Social Research at the University of Michigan, Ann Arbor. It provides a wide range of resources on digital curation and data management with particular relevance to the social sciences.

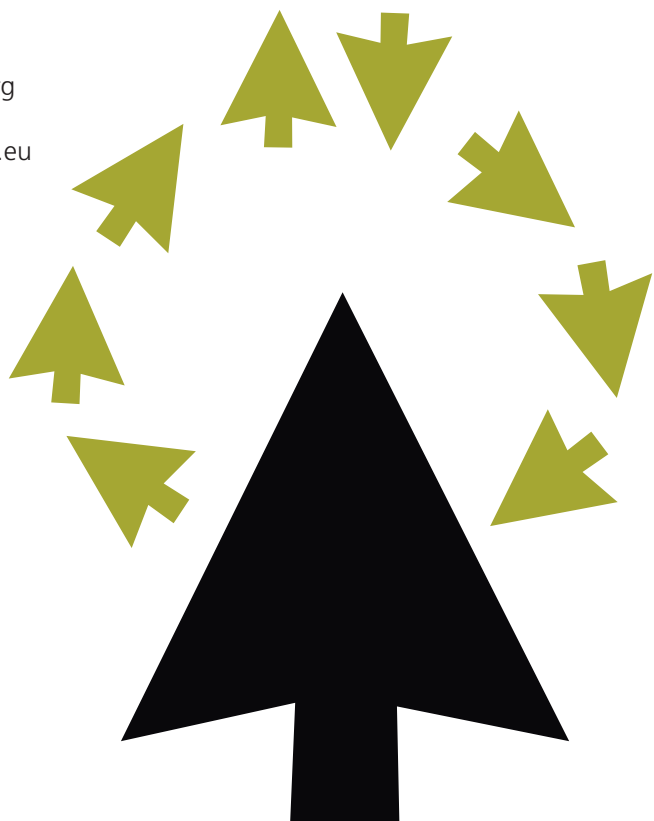
NESTOR — Network of Expertise in Long-term Storage of Digital Resources

NESTOR is a German network of expertise established to give guidance on digital preservation. It has developed criteria for a trusted digital repositories and handbooks and guides (mostly in German) addressing issues of digital preservation. The initial six year project (2003–2009) has been maintained as a cooperative network of the initial partners.

Membership Organisations

A number of international and national organisations aggregate and deliver domain-specific expertise on various aspects of digital preservation. A selection is outlined below.

- **ALLEA**, the federation of All European Academies, and the ALLEA E-Humanities Working Group – www.allea.org
- **Ariadne** – <http://www.ariadne-eu.org>
- **British and Irish Sound Archives** (BISA) – <http://www.bisa-web.org>
- **Coalition for Networked Information** (CNI) – <http://www.cni.org/about-cni>
- **Digital Library Federation** (DLF) – <http://www.diglib.org>
- **Digital Preservation Coalition** (DPC) – <http://www.dpconline.org>
- **Digital Production Partnership** – <http://www.digitalproductionpartnership.co.uk>
- **FIAT/IFTA** – www.fiatifta.org
- **IASSIST** – International Association for Social Science Information Services and Technology – www.iassistdata.org
- **IcarusNet** – <http://icarusweb.arhiv.hr>
- **International Association of Sound and Audiovisual Archives** – <http://www.iasa-web.org>
- **Open Planets Foundation** – <http://www.openplanetsfoundation.org>
- **PrestoCentre** – www.PrestoCentre.eu
- **Research Data Alliance** – <http://rd-alliance.org>
- **TEI – Text Encoding Initiative** – <http://www.tei-c.org/index.xml>



ARIADNE

ARIADNE is a not-for profit foundation that aims to facilitate the sharing and re-use of digital learning resources. Originally a EU project, it has now developed a repository, catalogue and validation services and provides guidance on standards and specifications for learning objects.

British and Irish Sound Archives

The BISA draws its members from institutions and individuals that hold sound or audio-visual collections and holds conferences and training events to assist those engaged in audiovisual archiving.

Coalition for Networked Information

The CNI is a US-based membership organisation comprising US libraries, publishers, academic institutions, IT companies and government agencies. Its focus is on IT development to facilitate scholarship and learning.

Digital Library Federation

The DLF draws its membership from mainly US-based organisations engaged in building digital libraries. It promotes work on 'digital library standards and best practice, research data management, aggregation and preservation services and resources which expand access to teaching resources for research, teaching and learning'.⁸

Digital Preservation Partnership

The DPC is a UK-based membership organisation focusing on advocacy and sharing expertise on digital preservation. Its members are drawn mainly from UK universities, libraries and archives. The Digital Repository of Ireland (DRI) and Trinity College Dublin Library are associate members.

Digital Production Partnership

The DPP is a partnership of UK-based public television broadcasters that have come together to develop standardised responses to emerging digital content.

The International Federation of Television Archives

FIAT/IFTA is an international professional association for audiovisual archives, particularly those in the television sector, as well as all the major national audiovisual archive bodies.

⁸ This quote is from <http://www.diglib.org/about> (accessed 16 May 2013).

International Association for Social Science Information Services and Technology

IASSIST is an international organisation of professionals working in and with information technology and data services to support research and teaching in the social sciences.

Icarus: International Centre for Archival Research

Icarus is a membership organisation based in Austria that draws its members from a network of archives. It runs workshops and educational seminars, and is a partner in a number of archival projects. As part of the European Network on Archival Cooperation (ENArC-project) it has developed ICARUS.net, an archival portal system that enables permanent online access to all data regarding archival records kept in different institutions that are members of the ICARUS network – archives, institutes and universities – all across Europe. It also participates in APEx (Archives Portal Europe), which aims at establishing an internet portal for documents and archives in Europe and additionally functions as the archives aggregator for Europeana.

International Association of Sound and Audiovisual Archives

IASA, based in Amsterdam, has members in over 70 countries and provides guidelines on best practice with respect to sound and audiovisual data.

Open Planets Foundation

The Open Planets Foundation is a network focusing on digital preservation tools and services that grew out of the EU Planets project.

PrestoCentre

The Presto Centre is a membership organisation based at the Institute for Sound and Vision in the Netherlands. It provides guidance and expertise in the field of audiovisual digital preservation.

Research Data Alliance

The Research Data Alliance is an emerging organisation (officially launched in March 2012) which has received funding from the European Commission through the iCordi project, the Australian Government through the Australian Data Service and the US National Science Foundation (NSF). They describe their goals as to “accelerate international data-driven innovation and discovery by facilitating research data sharing and exchange, use and re-use, standards harmonisation, and discoverability”.

Text Encoding Initiative

The TEI is a consortium that collectively develops and maintains a standard for the representation of texts in digital form. The original initiative was established in 1987; it became a consortium in 1999, and now has 77 institutional members.

EU Infrastructure Projects

In 2009, the EU instituted a legal framework designed to remove legal barriers to establishing multinational research infrastructure in Europe. This framework is known as the European Research Infrastructure Consortium (ERIC). Most current EU research infrastructure projects are now transitioning to ERIC status.

- **CESSDA**, Council of European Social Science Data Archives, an umbrella organisation for European social science data archives that co-ordinates national access to social science archives and datasets (preparing application for ERIC status).
- **DARIAH**, Digital Research Infrastructure for the Arts and Humanities, Digital Humanities (applied for ERIC status in October 2012) – <http://www.dariah.eu>
- **CLARIN**, Common Language Resources and Technical Infrastructure, EU language research infrastructure (obtained ERIC status in February 2012) – <http://www.clarin.eu/external>
- **ESS**, European Social Survey – multinational survey of attitudes, beliefs and behaviours (applied for the ERIC status in March 2012) – <http://www.europeansocialsurvey.org>
- **SHARE**, Survey of Health, Aging, and Retirement in Europe – multinational research databank on population ageing (obtained ERIC status in March 2011) – <http://www.share-project.org>
- **EGI**, European Grid Infrastructure (undertaking ERIC discussion process) – <http://www.egi.eu/infrastructure/index.html>

Digital Projects: Ongoing

Much of the cutting edge of digital repository development originates in projects funded by national governments, the European Union or private foundations such as the Andrew W. Mellon Foundation. It is not possible, due to resource limitations, to document the full depth and breadth of scholarship in this area. For example, in the period from 2005–2008 the EU funded over 60 digital repository projects under the eContentplus call. The projects here have been selected both to give a sense of the range of ongoing funded digital projects being developed in a variety of disciplinary fields and to identify projects with Irish participation.

- **APARSEN/APA** (the Alliance on Permanent Access to the Records of Science in Europe Network). The project aims to develop of sustainable European digital information infrastructure (due to complete in 2014). The Digital Preservation Coalition (of which DRI is a member) is a member of the network – <http://www.alliancepermanentaccess.org>
- **CENDARI** (Collaborative European Digital Archive Infrastructure) (due to complete in 2016). The project aims to integrate digital archives for medieval and modern European history. It is led by Trinity College Dublin – <http://www.cendari.eu>
- **CULTURA** (Cultivating Understanding Through Research and Adaptivity) (due to complete in 2014). The project will ‘deliver innovative adaptive services and an interactive user environment which dynamically tailors the investigation, comprehension and enrichment of digital humanities artefacts and collections’.⁹ Partners include Trinity College Dublin – <http://www.cultura-strep.eu/home>
- **DASISH** (due to complete in 2014). The project aims to explore possibilities for cross-fertilisation of existing social science and humanities research infrastructures. Partners include National University of Ireland Maynooth – <http://dasish.eu>
- **DECIPHER** (due to complete in 2013). The goal of this project is to ‘research and develop a reasoning engine, virtual environment and interfaces that can present

⁹ The quote is from http://www.cultura-strep.eu/home/cultura_vision (accessed 16 May 2013).

digital heritage objects as part of a coherent narrative, which is directly related to individual information searches and user contexts'.¹⁰ Partners include Dublin Institute of Technology, the National Gallery of Ireland and the Irish Museum of Modern Art.

- **EUDAT** (due to complete in 2013). The project aims to develop a collaborative research infrastructure – <http://www.eudat.eu>
- **EHRI**, European Holocaust Research Infrastructure (due to complete in 2014). The project aims to develop a portal to dispersed sources relating to the Holocaust – <http://www.ehri-project.eu>
- **LARM Audio Research Archive** (due to complete in 2013). The project aims to establish a national Danish audiovisual archive (similar to DRI) – <http://www.larm-archive.org>
- **NEDIMAH**, Network for Digital Methods in the Arts and Humanities (due to complete in 2015). The project examines the use of ICT methods in the arts and humanities. Partners included the Irish Research Council – www.nedimah.eu
- **OpenAIRE** and **OpenAIRE Plus** (due to complete in 2014). The project aims to create a database of EU research outputs – <http://www.openaire.eu>

¹⁰ The quote is from <http://decipher-research.eu/objectives> (accessed 16 April 2013).

Digital Projects: Completed

These projects addressed different aspects of the digital lifecycle, from developing a database of available resources or search portal (NISE, DRIVER, APEX) to considering issues of metadata classification and intellectual property rights (MILE). The DigCurV network, of which TCD is a contributing partner, established a curriculum framework from which vocational training programmes in digital curation can be developed. E-InfraNet, of which the Irish Higher Education Authority (HEA) is a partner, was tasked with co-ordinating national e-infrastructure policies and protocols, while DC-NET developed stronger coordination of public research projects in digital cultural heritage. The HEA was also a partner in SIM4RDM, which aimed to develop research data management knowledge, skills and support infrastructures. Many of the projects below, on completion of their core objectives, have persisted as networks of expertise.

- **APEX** – Archives Portal Europe (completed in 2012) – <http://www.archivesportaleurope.net>
- **CASPAR** – Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval (completed in 2007) – <http://www.casparpreserves.eu>
- **DC-NET** Digital Cultural Heritage NETWORK (completed in 2012) – <http://www.dc-net.org/>
- **DigCurv** – Digital Curator Vocational Education Europe) – (completed in 2013) – <http://www.digcur-education.org>
- **Digital Preservation Europe** (completed in 2009) – <http://www.digitalpreservationeurope.eu>
- **DELOS** – Network of Excellence on Digital Libraries (completed in 2009) – <http://www.delos.info>
- **Digital Library Interoperability**, Best Practices and Modelling Foundations (completed in 2010) – <http://www.dlorg.eu>

- **DRIVER** – Digital Repository Infrastructure Vision for European Research (completed in 2009) – <http://www.driver-repository.eu>
- **e-InfraNet** (completed in 2012) – <http://e-infranet.eu>
- **GRID2020** – Global Research Data Infrastructures (completed in 2012) – <http://www.grdi2020.eu>
- **MILE** – Metadata Image Library Exploitation (completed in 2009) – <http://www.mile.ssl.co.uk>
- **NISE** – National Movements & Intermediary Structures in Europe – <http://nise.eu/en/project>
- **Planets** – Preservation and Long-term Access Through Networked Services (completed in 2010) – <http://www.planets-project.eu>
- **Project Bamboo** (Phase 1 completed in 2012) – www.projectbamboo.org
- **Research Space** (completed in 2012) – development of tools to assist in annotation and sharing – <http://www.researchspace.org>
- **SIM4RDM** – Support Infrastructure Models for Research Data Management (completed in 2013) – <http://www.sim4rdm.eu>
- **Shaman** – Sustaining Heritage Access through Multivalent Archiving (completed in 2007) – <http://shaman-ip.eu>

Conclusion

The UNESCO Charter on the Preservation of Digital Heritage stated in 2003 that ‘digital heritage is at risk of being lost and that its preservation for the benefit of present and future generations is an urgent issue of worldwide concern’.¹¹ DRI’s mission is the digital preservation of and access to Ireland’s cultural and social heritage for present and future generations. Mapping international approaches to caring for digital content is an important information-gathering exercise, which we are happy to share here with the community, in the hope that examples of best practice can inform and shape our community’s advancement as a whole.

Preservation of our cultural and social heritage is an important societal responsibility commonly associated with galleries, libraries, archives and museums. Digital repositories have emerged relatively recently (see Figure 1). Among the earliest to be established were MERLOT and the Internet Archive in 2000. The mapping exercise we have carried out has shown that it is possible to classify three different architectural approaches to caring for digital content:

1. **Single-site repositories**, in which the technical and organisational function are located in one place (excluding off-site backup). The single-site approach is often adopted by national infrastructures.
2. In 2007–2009 a number of **metadata aggregators** were established. This approach brings together (aggregates) the metadata of a number of single-site repositories, thus increasing user awareness of content held in various repositories.
3. Since 2009 there has been a demonstrated shift towards the establishment of **multi-site repositories**, in which the technical infrastructure is federated across a number of repository sites. The Internet Archive and Dataverse were early adopters of such a multi-site approach.

In seeking to future-proof the DRI infrastructure, and in line with this emerging trend, we have adopted a federated architectural approach for the DRI. This also enables us to partner with existing and future digital archives, which we view as essential for a richer user experience, and to truly achieve a national mandate.

¹¹ This quote is from http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html (accessed 30 April 2013).

Another observation arising from our mapping is the different emphases that digital projects place on access to versus preservation of the digital content. Current funding trends arguably appear to favour access, visualisation, user engagement tools and mechanisms, rather than digitisation, sustainable data management, and digital preservation. DRI urges some caution here: while user access and engagement are absolutely essential, it is equally important to consider the longer-term responsibility to preserve this rich digital content, so that future generations will also have the opportunity to access and engage with it. As Steve Knight of the National Library of New Zealand noted in 2012, in reference to the UNESCO statement and similar documents, ‘the promise of these documents has not yet been met and digital preservation practice still falls far short of the objectives that their authors envisaged even though the potential to collect, preserve and make accessible a fuller expression of the cultures, heritages, histories of peoples is within our grasp’.¹²

A further trend we identify and applaud is the increasing emphasis placed on policy engagement and training and education programmes. Digital preservation and open access form a digital ecosystem, and for a system to work effectively all aspects must be addressed and sustained. Organisations such as the Digital Curation Centre (DCC) and the Digital Preservation Coalition (DPC), which inform policies and provide training and education programmes in digital preservation and curation, are an important part of the data ecosystem. Global organisations such as the Research Data Alliance that seek to influence practice and culture of researchers and data practitioners will have an important long-term positive impact on data preservation.

Digital preservation and access to our cultural and social heritage are challenging goals but absolutely essential. The mapping exercise that we report here is not exhaustive and represents a snapshot in time. DRI will continue to monitor international developments and identify emerging practices, projects and organisations, and in turn we will contribute to the national and international landscape. DRI is actively engaging with international networks and projects in order to share best practices with the community to enable cost savings and improved standards of preservation and access, for the collective good.

¹² This quote is from the Keynote presentation to the iPres Conference, Toronto, Canada, 5 October 2012.

Figure 1: Year of Origin of Repositories

