

Digital Archiving in Ireland

National Survey of the
Humanities and Social Sciences



Digital Archiving in Ireland

National Survey of the Humanities and Social Sciences

Aileen O'Carroll and Sharon Webb

First published in 2012 by the National University of Ireland
Maynooth, Maynooth, Co Kildare.

© National University of Ireland Maynooth

When citing this report, please use the following wording:

O'Carroll, A. and Webb, S. (2012), *Digital archiving in Ireland: national survey of the humanities and social sciences*. National University of Ireland Maynooth. DOI: 10.3318/DRI.2012.1

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

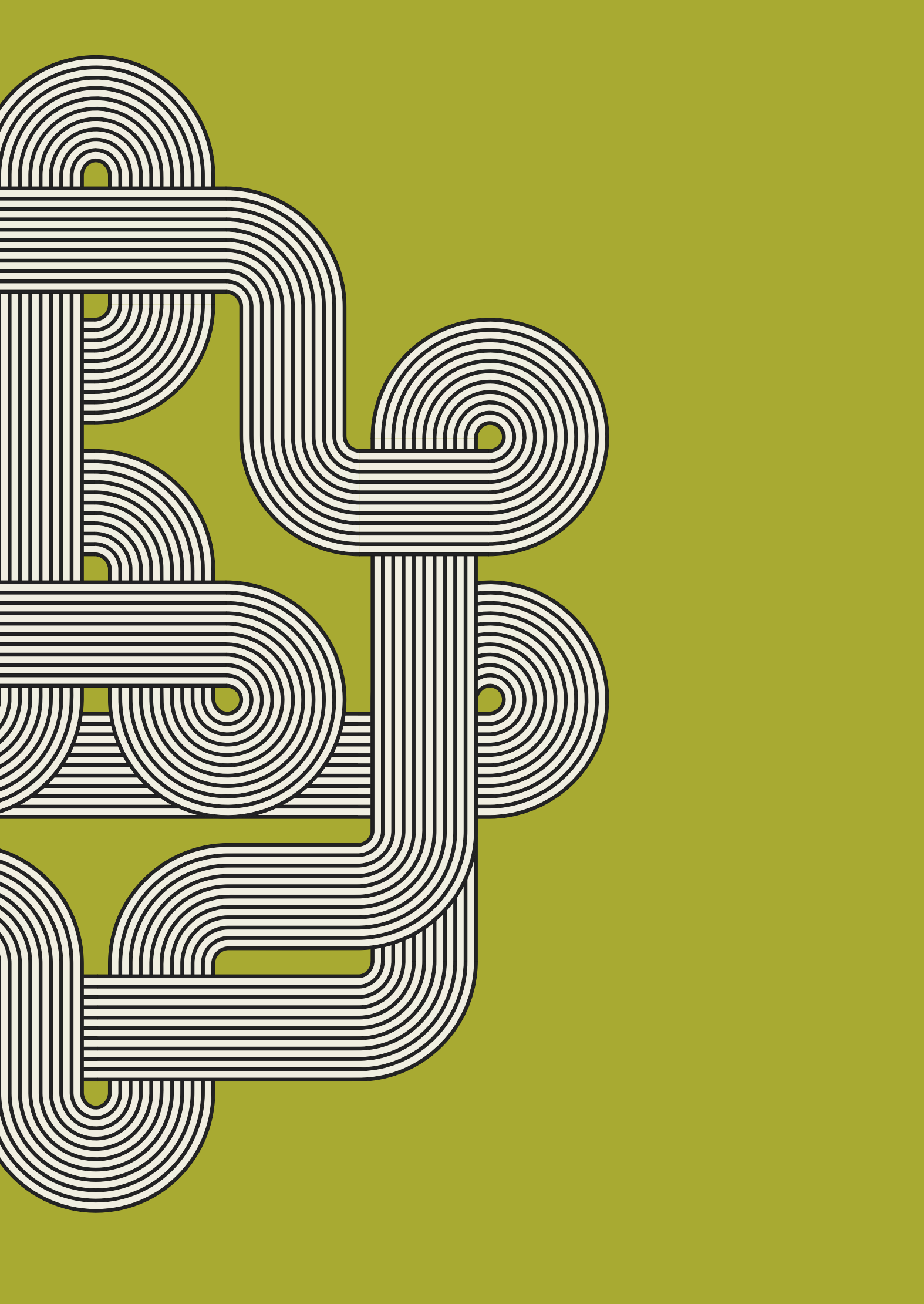
ISBN 978-0-956326-76-8

Design and layout by Fidelma Slattery

Printed in Ireland by Walsh Colour Print

Contents

5	Director's foreword
6	The Digital Repository of Ireland
7	Executive summary
10	Introduction
12	Methodology
15	Types of digital data in Ireland
18	Digital preservation: 'sustainable access'
27	Digital file formats
32	Metadata and vocabularies: describing data
37	User tools
44	Structuring content: database formats/systems, content management systems and repository software
47	Absences
48	Conclusion
51	Appendix 1: Methodology
60	Appendix 2: Resources
64	Appendix 3: Stakeholder Advisory Group members



Director's foreword

The Digital Repository of Ireland (DRI) is conducting a national programme of stakeholder interviews to determine the digital preservation and access practices in cultural institutions, libraries, higher-education institutions, funding agencies and more. Our findings shape our requirements specification in building the national repository, and they are also the beginning of a process to agree national guidelines on digital preservation for the humanities



DR SANDRA COLLINS

and social sciences. Our approach is first to determine national practice, then to work with the community in building national guidelines and hence to inform national policy.

This report presents our first findings. It is important to share our experiences, to learn from one another and from best practice both nationally and internationally, in order to serve our community of users now and into the future. Community engagement and informed dialogue are an essential part of this.

I would like to convey my sincere thanks both to the authors of this report and to the 40 respondent institutions that gave their time and support so generously towards this goal.

Much work remains to be done; further engagement is under way with our wide circle of stakeholders, and ascertaining digital practices is only the first step towards national guidelines that will be designed with the community, to be used by the community. It is a challenging task and not one that we undertake lightly, but it is essential and we must act now.

The DRI is working to raise awareness of the need for and benefits of digital preservation and open access, while respecting and acknowledging ownership, copyright, intellectual property rights, privacy and confidentiality. Digital preservation of our social and cultural heritage is imperative, and this is exactly what is at stake today, unless we act together.

Dr Sandra Collins

Director, Digital Repository of Ireland
Royal Irish Academy

Digital Repository of Ireland

The Digital Repository of Ireland (DRI) is building an interactive national trusted digital repository for contemporary and historical, social and cultural data held by Irish institutions. The DRI is linking together and preserving the rich data held by Irish institutions, providing a central internet access point and interactive multimedia tools for use by the public, students and scholars. The DRI is a national e-infrastructure for the future of education and research in the humanities and social sciences.

The DRI Research Consortium comprises the partners: the Royal Irish Academy (lead institute); the National University of Ireland Maynooth; Trinity College Dublin; the Dublin Institute of Technology; the National University of Ireland, Galway; and the National College of Art and Design. We are also collaborating with a network of academic, cultural, social and industry partners, including the National Library of Ireland, the National Archives of Ireland and Raidió Teilifís Éireann. We were awarded €5.2m from the Higher Education Authority's Programme for Research in Third-Level Institutions, Cycle 5 (funded as the 'National Audio Visual Repository'), and have also received awards from Science Foundation Ireland, Enterprise Ireland and the Ireland Funds.

...an interactive national trusted digital repository for contemporary and historical, social and cultural data held by Irish institutions.

Please visit our website, www.dri.ie, to learn more.



Executive summary



DR AILEEN O'CARROLL



DR SHARON WEBB

The Digital Repository of Ireland interviewed 40 institutions concerned with the humanities and social sciences about the procedures and practices that they have adopted in order to archive and care for the data in their collections. The interviews focused in particular on the care of digital data. The DRI will use the information generously provided to inform both the design and implementation of the national repository and the development of national guidelines, which will be designed with the community, for use by the community.

Our findings to date address different aspects of the digital lifecycle and are summarised below.

Types of digital data

A wide range of types of digital data were being cared for and created, including digitised manuscripts, photographs, moving images and audio material, as well as geospatial and geographical raw data. Digital data were generated by the digitisation of analogue material in collections, but increasingly data are created in digital form, that is, they are 'born-digital'. Most social scientific data, academic outputs and organisational data (including minutes of meetings, e-mails, webpages and social media) are now born-digital.

Sharing and reuse

There was an eagerness to enable sharing and reuse of digital data. Some collections, however, had copyright or ethical restrictions that limited these possibilities. There is a need for a national policy that would enable increased sharing and reuse of digital data.

Preservation

The preservation of digital objects was identified as a key challenge, a challenge that in many ways is more complex than the preservation of analogue objects. Digital preservation requires not only the secure storage of digital materials but also policies and workflows that ensure that such materials will be accessible and usable in the future. Although many interviewees were meeting the challenges of secure storage, few had workflows in place to ensure the successful migration of objects as current formats become redundant. Many of the interviews identified skill shortages and a lack of appropriate technical infrastructure as a key barrier to ensuring long-term preservation of digital objects. Born-digital data are in most danger of being lost to future generations.

Storage and formats

Although many institutions were able to store their current digital data, there were fears that, as the size of digital collections continued to grow, it would become difficult to afford and manage the storage space necessary. This was particularly the case in fields that cared for audio-visual files and 3D modelling files.

Data were stored in a relatively small number of formats. Institutions distinguished between preservation formats and access formats, often creating digital objects in both. Some formats that were appropriate to the data and in use internationally were not widely evident in Ireland. Policy guidance is required for those data types without archival formats in place and for format migration.

Metadata and interoperability

The added value of digital data over analogue data is that they enable institutions to share and build connections between collections. This is only possible, however, if standard metadata, fixed-word vocabularies and ontologies are used. Most of the metadata standards used were appropriate to the type of data that they were applied to; however, the DRI will face a critical and difficult challenge in ensuring that it is possible to build connections between datasets that are created using a range of metadata standards.

User tools

There is an increasing desire to ensure that today's users are able to interact with and add value to digital data. Many of the interviewees provided enhanced access either through curated collections and the provision of user tools, in particular geospatial mapping tools, which enabled users to manipulate data online, or through mobile applications that delivered content in new ways to users. A number used social media to raise the profile of objects in their collection and to generate new information about

the objects. These are welcome developments; however, they also raise the challenge of ensuring the sustainability of these user tools and the incorporation of user-generated content into existing collections.

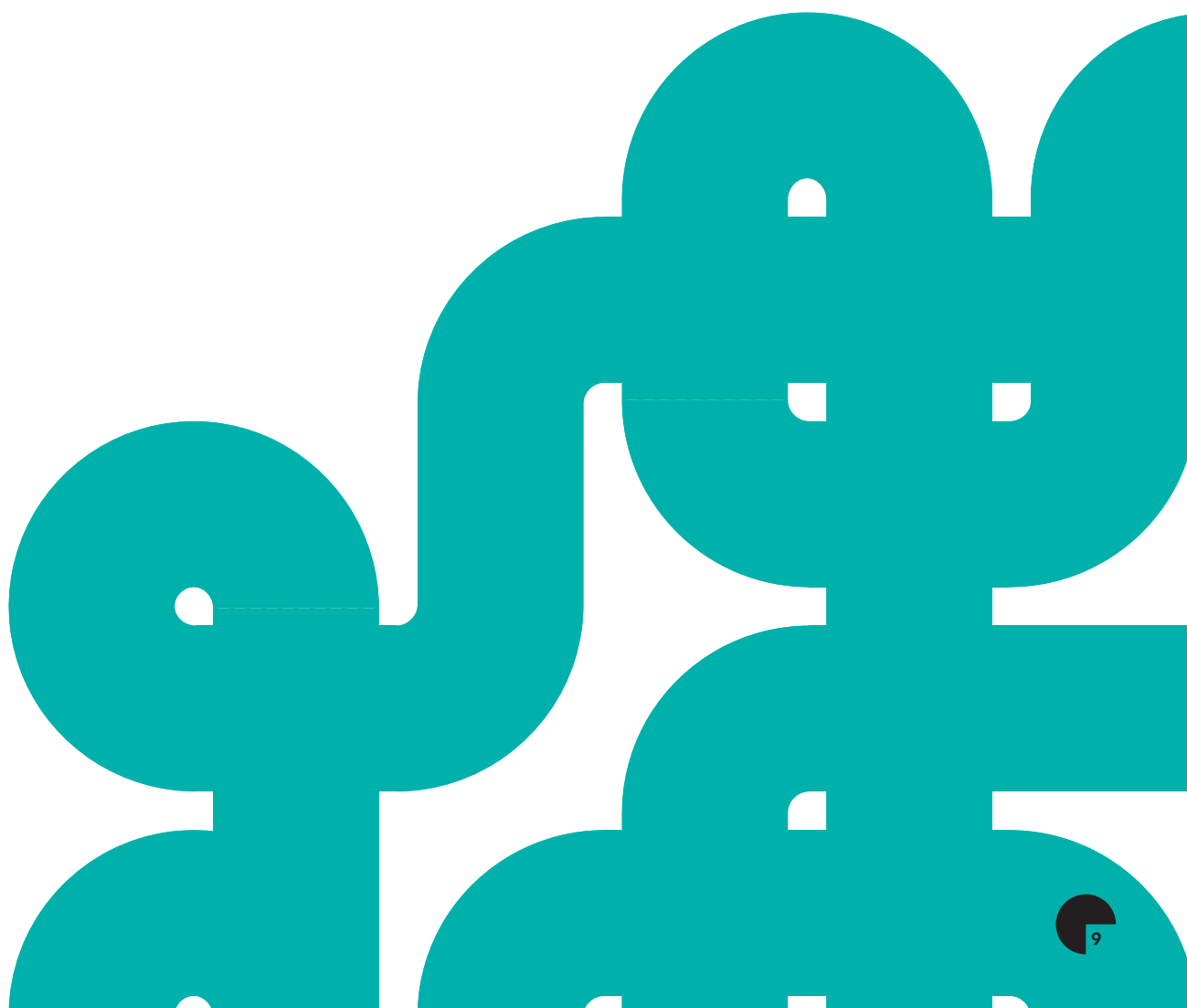
Structuring content

There are many database, repository and content management systems available. While MySQL was clearly the most popular database in use, there was surprisingly little consensus on the content management systems used, reflecting both the diverse needs of the community and the crowded nature of the content management field. Many institutions were upgrading or moving to new systems, in particular to ensure greater web access to their collections. This creates a challenge in integrating new and legacy systems.

Our findings reveal a vibrant and active community, which is cultivating and developing Ireland's digital landscape at a challenging time. The DRI will work with the community to develop digital guidelines and to provide preservation and access services to enhance current offerings.

Dr Aileen O'Carroll and Dr Sharon Webb

National University of Ireland Maynooth



Introduction

A trusted digital repository can be defined as a technical infrastructure ‘whose mission is to provide reliable, long-term access to managed digital resources for its designated community, now and in the future’.¹ The primary aim of the Digital Repository of Ireland (DRI) is to develop such an infrastructure. Yet, while we focus on the technical development of this system, a key issue is providing the associated services to the DRI’s ‘designated community’. This ‘designated community’ is diverse and is represented in the range of stakeholders with which the DRI has engaged to date. The successful implementation of the DRI’s goals and deliverables depends on the system’s ability to satisfy and implement the requirements of the DRI’s community. As part of this process, three members of the DRI research team, Dr Sandra Collins (Director), Dr Aileen O’Carroll (Policy Manager) and Dr Sharon Webb (Requirements Manager), carried out stakeholder interviews that surveyed the community’s current activities, requirements and desires in terms of digitisation, digital asset management, digital preservation and user engagement, as well as the challenges faced by, and the opportunities open to, this community.

These interviews are essential to the DRI’s ability to deliver a system that caters for the needs of its users. They inform the development of systems requirements and the DRI’s policy and usage guidelines. By incorporating requirements interviews into policy formulation and management, we aim to address the key concerns of the community and to develop strategies for digital rights management, digital preservation, access control and digital standards in response to those stated needs. The primary objective of these interviews is to ensure that the system is informed by authentic user requirements and that, as much as is possible, we support current good practices (e.g., data formats and metadata standards). This will allow us to build on the experience

The primary objective of these interviews is to ensure that the system is informed by authentic user requirements and that, as much as is possible, we support current good practices.

¹ *Trusted digital repositories: attributes and responsibilities*, an RLG–OCLC report (2002), available at <http://www.oclc.org/resources/research/activities/trustedrep/repositories.pdf> (accessed on 21 August 2012).

of the community in providing preservation and access services, while adding value and innovation to humanities and social science data through the DRI's infrastructure.

This report presents our findings from the initial phase of interviews and reveals that, although there is a diversity of interests and practices stemming from various perspectives, backgrounds and institutional obligations and remits, members of the DRI's designated community face many of the same issues and challenges in securing Ireland's digital cultural and social heritage. The report reveals a vibrant and active community that is cultivating and developing Ireland's digital landscape and provides essential information that will directly inform the DRI's development. This community is proactively unlocking Irish archives, providing access to content in new ways that add value to the material and transforming the user's experience.

This report also provides an opportunity for our stakeholders and interviewees to engage with this community. Along with the interviews, this report represents initial steps in terms of the DRI's stakeholder engagement. This dialogue will continue.

It is important that we acknowledge that this research builds on and complements previous Irish publications in this area, including the Library Council, *Our cultural heritage: building the gateway* (Dublin, 2004), the Irish Manuscripts Commission and the Digitisation Task Force, *Digitisation policy* (Dublin, 2007), the Spatial Heritage and Archaeology Research Environment IT, *A survey of digital practices in Irish archaeology* (Dublin, 2008), the Irish Qualitative Data Archive and the Tallaght West Childhood Development Initiative, *Best practice in archiving qualitative data* (Dublin, 2011), to name but a few. We would also like to acknowledge that this work builds on and complements recent and ongoing Irish initiatives. We envisage that this report will supplement that work and add to the enormous efforts already undertaken by our interviewees to develop and secure Ireland's digital future.

The DRI research team would like to thank everyone who has talked with us to date, and we look forward to engaging further with the community.

Methodology

The DRI conducted 40 requirements interviews with key stakeholders from December 2011 to August 2012. Figure 1 indicates the main spheres from which the interviewees were drawn and illustrates the range of stakeholders with which the DRI must interact, in terms of those who currently hold or produce digital content and those who use that content (there is, of course, overlap between the two categories; e.g., an academic researcher may both create research data and use data created by others). Individuals from the groups marked with an asterisk have been interviewed to date. The interview process will continue, to include those groups not interviewed in the first phase of consultation (see Appendix 1 for a list of organisations consulted).

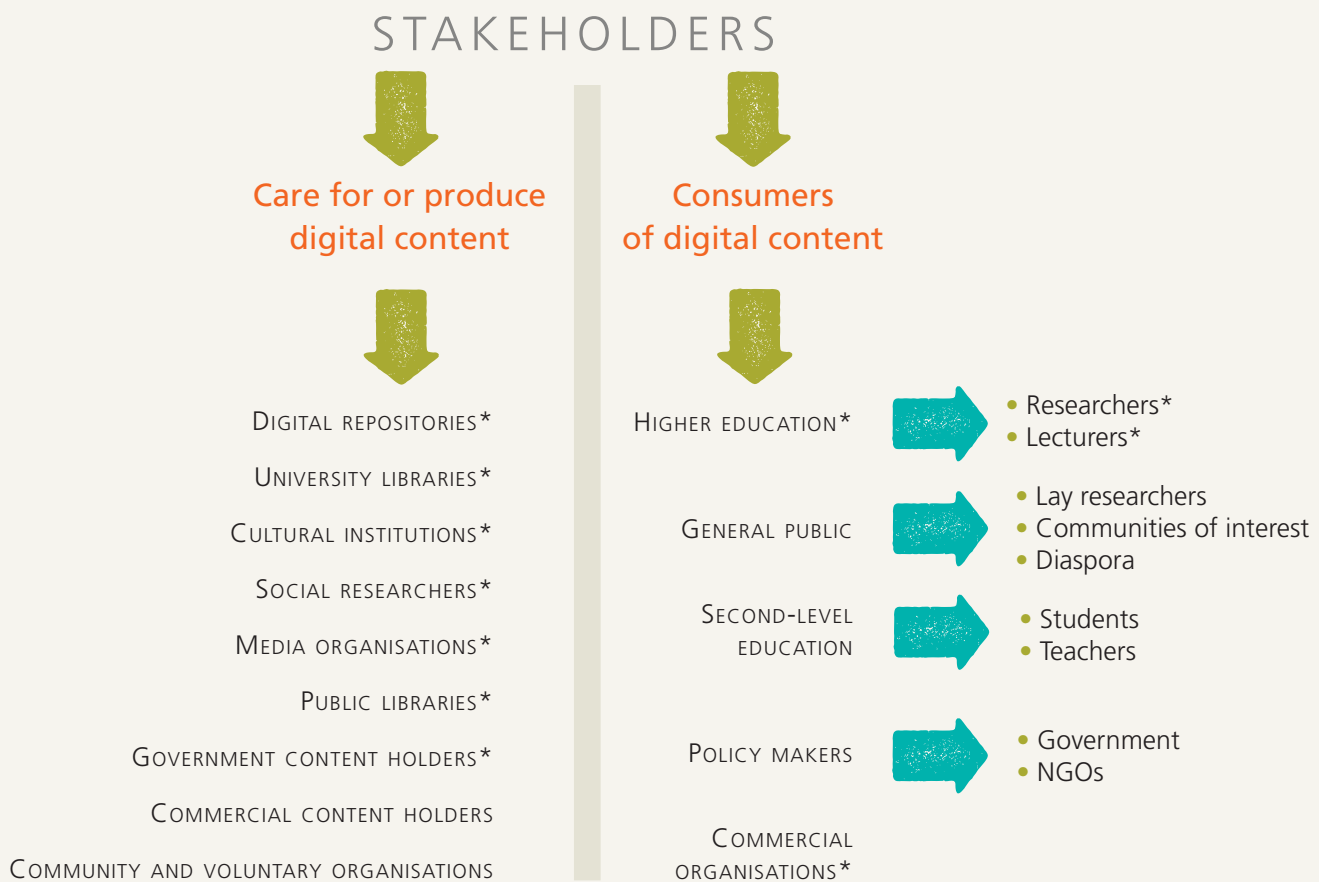


Fig. 1: Stakeholder interviews

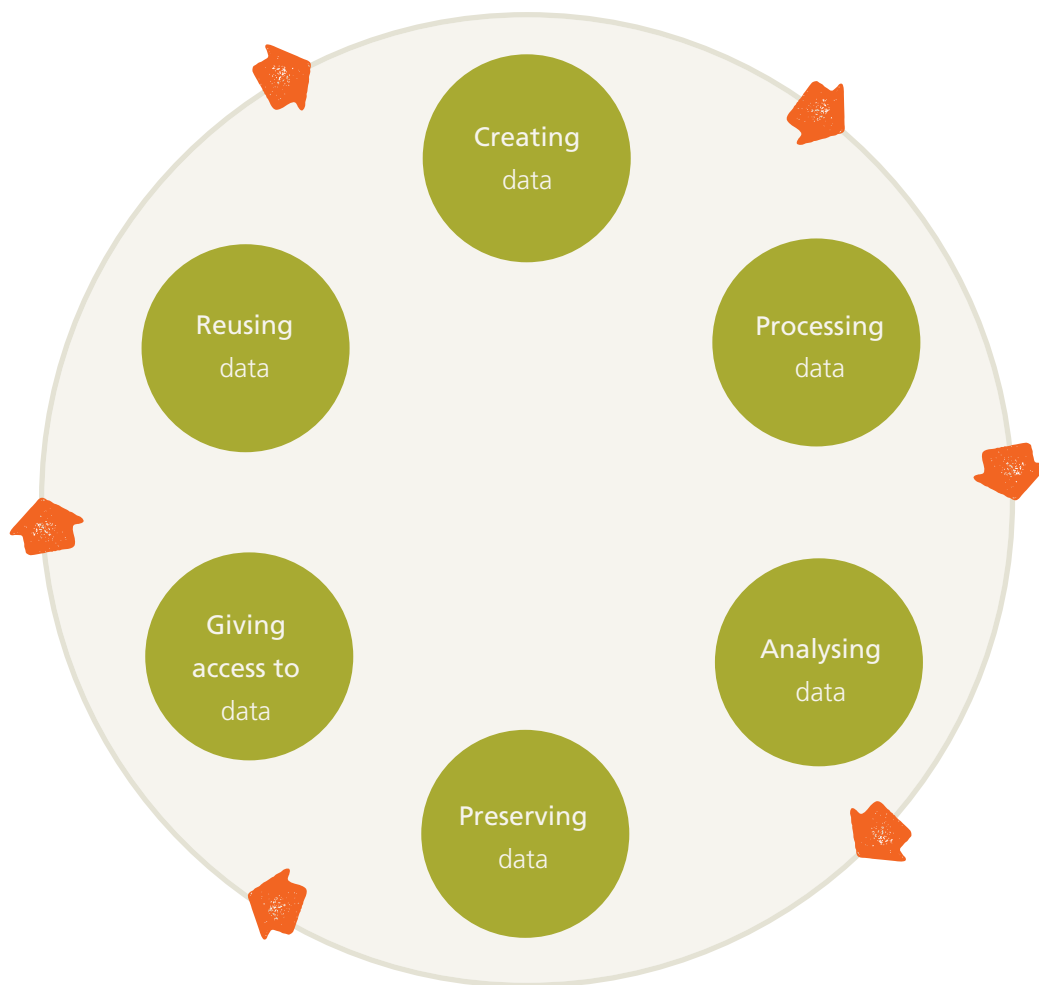


Fig. 2: Representation of the data lifecycle produced by the UK Data Archive²

'An inquiry, not an inquisition'³

Stakeholder and user interviews were structured but semi-formal, using a topic guide approach that allowed the interview to unfold as a free-flowing conversation while ensuring that all of the designated topics were discussed at some point. The topic guide ensured that the DRI interviewers received the desired information on a range of issues pertinent to requirements analysis and policy development, including data formats, metadata standards, existing systems, approaches to future challenges and expectations. We used the data lifecycle model developed by the UK Data Archive, detailing the phases of data creation, processing, analysing, preserving, access and reuse, as an initial topic guide template for the stakeholder interviews (Fig. 2). Pilot interviews were conducted with DRI partner organisations; the process was refined; and a final topic guide was developed that was used to inform further interviews (see Appendix 1).

² 'The Data lifecycle', UK Data Archive available at www.data-archive.ac.uk (accessed on 5 October 2012).

³ Wieggers, 2006, p. 57.

Ethics approval

The interviews were recorded (audio only) with the interviewees' permission. It is planned that these interviews will form a collection within the DRI and become part of the repository's project history. The interviews capture a moment in time after the initial phase of Ireland's digital archiving and before the wider, systemic approach being attempted through the DRI and therefore provide a view of Ireland's digital landscape at a critical juncture, highlighting the areas that are flourishing and those that are in need of support.

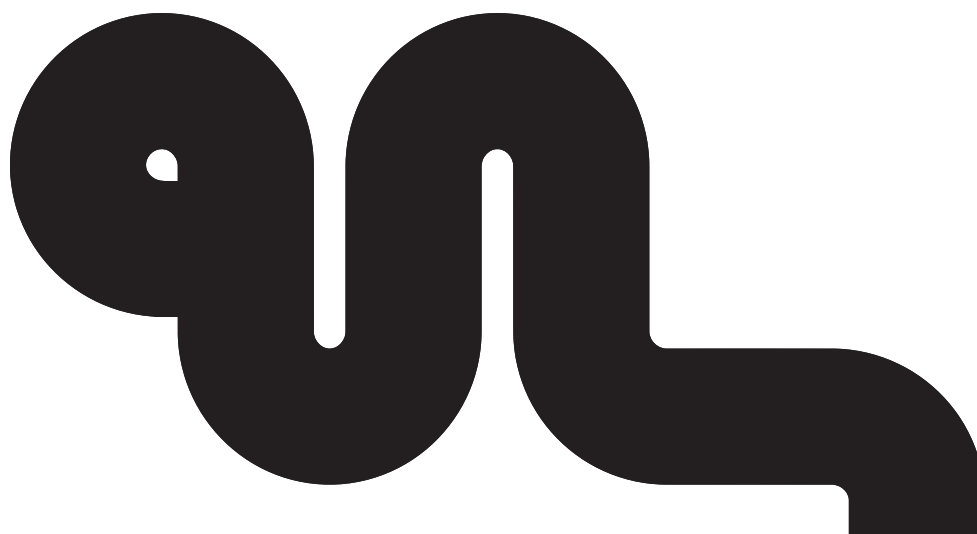
A copy of the consent form is provided in Appendix 1. Some interviewees requested that they not be identified, and consequently their responses are anonymised and unattributed in the text below.

Twenty-seven of the interviews have been transcribed and coded using the qualitative data analysis software MaxQDA. Encrypted WAV, RTF and Microsoft Word files are stored locally with the research team, and a back-up copy is stored unencrypted in a fireproof safe in a swipe-card-protected laboratory at the National University of Ireland Maynooth (NUI Maynooth). These files are accompanied by metadata using Dublin Core and Open Archives Initiative standards (see http://www.openarchives.org/OAI/2.0/oai_dc.xsd).

The interviews...provide a view of Ireland's digital landscape at a critical juncture, highlighting the areas that are flourishing and those that are in need of support.

Review process

This report was submitted for review to the DRI partners. After approval, it was submitted to the DRI Stakeholder Advisory Group (see Appendix 3). Feedback from both groups has been incorporated into this final report. Feedback included the suggestion that the DRI would conduct an audit of digitisation resources among the community to facilitate sharing of said resources. Additionally, a respondent indicated that a further survey to obtain more detailed information on the scale of data held by institutions and on technical infrastructures would be beneficial.



Types of digital data in Ireland

Of the 40 institutions interviewed, 36 were responsible for digital data. Because of the wide range of institutions, their relationship to the data varied markedly. Some were data creators, and some had a role in holding data for use by others. Others provided services that assist in the reuse and analysis of data that were also held elsewhere. The majority of the organisations interviewed held both analogue and digital data, digital data being a smaller, but growing, proportion of their collections. In terms of the types of digital data held, the range outlined below mirrors the variety of institutional roles.

-  Electronic text documents (transcripts, essays, diaries, theses, journal articles, books, journals, reports, learning resources)
-  Digitised images of analogue manuscripts, paintings, archival documents, newspaper clippings, printed ephemera
-  Photographs (including born-digital and digitised copies of prints and film negatives), some of which are historical and some contemporary, created in the research process or to document installations and other artworks
-  Moving images (including material produced for broadcast, home videos, born-digital and digitised copies of film and VHS, some including subtitles)
-  Interview and focus group audio files, home recordings
-  Radio programmes (born-digital and digitised copies of analogue radio, cassette tapes, records and cylinders)
-  Webpages, e-mails, social media, podcasts
-  Geospatial surveys, 3D documentation of objects, maps, architectural plans

Fig. 3: Types of digital data held by stakeholder groups

In the vast majority of cases, the collections were continuing to be added to, that is, they can be considered to be 'living archives' that develop and grow over time. Twenty-six of those interviewed had Irish-language data, and sixteen had data in other languages (such as Latin, French and Spanish).

'Born-digital' refers to material created in digital form; for example, an e-mail is a born-digital object. Thirteen of the institutions had born-digital material as part of their collections. For some, such as radio and television broadcasters and social research institutions, the transition from analogue to digital data is almost complete, with future data being generated and stored in digital form. Other institutions, such as art galleries, museums, and national and cultural archives, hold considerable analogue data. Here, digital data emerge from the institutions' own digitising processes (where progress depends on limited or reducing resources and funds). With all groups, some types of data, such as correspondence, are increasingly being generated in digital form, as e-mail replaces the letter. Other types of digital data, such as social media data (Flickr, YouTube, Twitter and Facebook), are being used in social scientific and humanities research by many of the institutions interviewed, but there is no concerted strategy to ensure that these data will be archived for the future.

Many of the interviewees were committed to allowing public access to their data. As Siobhán Fitzpatrick, Librarian at the Royal Irish Academy, explained:

[it] is a very important thing for us...the idea of having materials freely accessible as far as is humanly possible...Our overall access policy is to make material as freely available as possible...Most of the things that we would have done were paid for out of the public purse and our salaries are also paid by the taxpayer.

Copyright and confidentiality restrictions, however, limit the ability of institutions to share data. Copyright issues were of concern to many. Libraries were affected by the impact of copyright legalisation that placed access restrictions on the books, journals and collections that they held. Some institutions exercised copyright to generate revenue. Others exercised their copyright in order to limit unwanted reuse of their data; for example, one institution cited the reuse of a photograph in its collection by a commercial entity in a way that exposed the individuals in the photograph to ridicule. This type of misuse could be prevented by denying the right to reuse, although this requires that the institution be both aware of the reuse and in a position to defend its copyright.

Most social scientific data (and some donations to libraries and archives) had reuse restrictions placed on them that limited who would be able to access the data and

required that the anonymity of the interviewees be maintained. These limitations lessen over time: in 100 years, all data can be shared. Interviewees expressed concern, however, about the implications of such limitations for long-term preservation. The time and resources needed to ensure sustainable access to these objects, in order for them to become publicly available in the far future, had not been fully explored by any of our interviewees. It is, however, an area that requires immediate attention, and a number of institutions are developing pilot projects to address it.

In conclusion, institutions tasked with the care of data from the humanities and social sciences hold digital data in a wide variety of forms, ranging from the textual to the audio, the visual and the moving image. Although, in many cases, analogue collections are far greater than digital collections, digital data are increasing. In particular, social scientific research data, publishing and audio-visual data are increasingly created only in digital form. Digital data cannot be thought of as a simple copy of a physical object, and this report identifies many ways in which digital data differ from analogue data. Digital data create new possibilities but also new challenges. One of the possibilities is that there is a greater opportunity to share and reuse the data, such that the public has greater access to data held and produced by publicly funded institutions. This review found a marked interest in increasing access to digital data, including the use by many institutions of social media to engage with the public. However, there are important tensions. In the social sciences, where data are collected on the lives of contemporary individuals, a balance needed to be maintained between the rights of the public to access publicly funded data and the rights of research participants to have their confidentiality protected. Copyright brought an additional set of tensions that both restricted the sharing of data and protected the interests of individuals and institutions. While the copyright concerns attached to digital and physical objects are in very many ways similar, digital data carry additional opportunities and challenges. It is much easier to make collections and objects widely available by sharing them on the internet, but there was a clear sense that, once an object is released, it is extremely difficult, if not impossible, to police how that object might be used. Given that we are living in an increasingly digital world, there is a need for a national digital policy that capitalises on the possibilities of digital data and provides guidance on how to facilitate their sharing and reuse.

Given that we are living in an increasingly digital world, there is a need for a national digital policy that capitalises on the possibilities of digital data...

Digital preservation: 'sustainable access'

A major shift in the nature of living archives was evident, especially for those who dealt with modern content rather than solely historical material. Only one of our interviewees described its archive as static, that is, the institution did not actively receive or seek new analogue objects or material. Yet, with regard to digital collections, all of our interviewees had growing archives, indicating that the long-term preservation of digital data, as well as their capture, is a significant issue.

Digital preservation is therefore a challenge for all stakeholder institutions. Preservation is concerned with providing long-term access to digital objects, preserving continuity in form as well as functionality. It is not simply a back-up of data, because long-term digital preservation must consider format, software and hardware obsolescence, among other issues. Although it is possible for anyone to read a page from a book written 100 years ago, the same is not true of a floppy disk containing WordPerfect files from twenty years ago. Preservation is also resource-intensive and expensive, and all of our interviewees faced the challenge of providing long-term commitments to digital preservation, given current resource and funding restrictions and curtailments. The majority of archives, libraries, museums, universities, research institutes and other content owners from both the public and the private sphere that were interviewed had encountered difficulties with the preservation of digital data.

Although it is possible for anyone to read a page from a book written 100 years ago, the same is not true of a floppy disk containing WordPerfect files from twenty years ago.

After digitisation: 'custodianship of digital data'

Many institutions digitise on an ad hoc basis, either to meet user demands or for particular projects, for example, online exhibitions. One reason for this selection strategy is a lack of funding to digitise more comprehensively. Most telling, however, is that some view the creation of digital surrogates, at present, as an unreliable method of long-term preservation and see microfilm as a superior, tried-and-tested method. Yet, we must be clear: digitisation is not preservation. Institutions and funding agents need to consider the 'custodianship of digital data' after the digitisation process is complete, a step of

data management that is often overlooked in project funding. A number of our interviewees identified this problem, stating that although funding was allocated to digitise content, no allocation was made for the ‘custodianship of [the] digital data’⁴ generated through a funded project. While money was, and still is, allocated to generate digital surrogates of analogue objects, the long-term or even medium-term preservation of these digital objects was not accounted for as part of funding streams. This indicates that there is a significant funding, as well as methodological, gap between the generation of digital objects and the long-term preservation of content on completion of a project. This problem was also identified in research projects that produce various datasets during their lifetime, after which there are no contingencies for the long-term deposit of data or data management plans that would facilitate archiving and reuse. This problem could lead to duplication of research effort and costs.

Another view on the problem of digitisation is that while some funding may be available to create digital surrogates, there is little allocation to stabilise or conserve the original artefact. A number of institutions felt strongly that digital surrogates enhance access but should not be viewed as replacements. Yet digital preservation and access or dissemination of content are not mutually exclusive. Hugh Murphy, Senior Librarian, John Paul II Library, NUI Maynooth, commented that digital preservation facilitates ‘the user of today — the researcher of today’ but must also ensure the same level of quality and access for ‘researchers in twenty or thirty years’ time’.⁵

Born-digital data: ‘The [data] in most danger’

Born-digital data and archives pose more of a preservation problem than paper-based or physical objects and collections. Born-digital data are more complex than analogue data, as they encompass a multitude of data types, formats, applications and operating systems, as well as other hardware and software requirements and dependencies. One archivist stated:

in our case it is the born-digital [content] that is the big issue...that is, the [data] in most danger...We have no way of preserving it right now...the born-digital [data] that is being created right now or has been created over the last 30 years in so far as [what] survives...is the huge problem. And it is a problem everywhere; it is not just for us. But that is the one where we need to be looking for solutions.⁶

⁴ Anthony Corns, The Discovery Programme.

⁵ Hugh Murphy, John Paul II Library, NUI Maynooth.

⁶ Anonymous institution.

The major difficulty with born-digital content is that it is practically impossible to preserve or save it retrospectively. Digital data and media are fragile and volatile. There is therefore an immediacy of effort required to preserve long-term access to digital content. Una Walker, a post-doctoral researcher at the National College of Art and Design and a member of the DRI, shared her concerns on preservation of born-digital material: 'born-digital works...present particular problems in relation to preservation, partly because some [could] be networked, some [could] be using social media'.⁷ This indicates how born-digital works, in this instance new media artworks, are more complex than the digital surrogates of analogue material. Dr Walker is also closely involved with the National Irish Visual Arts Library, which has created a pilot project, the Digital Ephemera Archive, promoting the capture and preservation of digital ephemera.

The major difficulty with born-digital content is that it is practically impossible to preserve or save it retrospectively.

A small number of institutions informed us of pilot projects to capture web content based on a special remit such as a particular event or subject or to complement paper-based collections. The National Library of Ireland's born-digital collection includes 'web archiving activities around the 2011 General Election', hosted by the Internet Memory Foundation.⁸ The National Library of Ireland's website states that 'it is working towards collecting other born-digital material'.⁹ However, of the institutions and archives to whom we spoke, none indicated that it plans to web archive on a continuous or encompassing basis. Rather, web archiving is viewed as an activity related to particular projects and as such is restricted by limited resources and funding. As a result, very few web collections are being generated or maintained. However, organisations such as the Internet Archive, which offers services such as Archive-It, 'allow[ing] institutions to build and preserve their web archive of digital content',¹⁰ and the Internet Memory Foundation, as used by the National Library of Ireland, provide solutions to harness and gather institutional web content and could also be used to generate and host national web-based ephemera, material, data and collections. A limitation of these services is that they are hosted abroad, so that Irish institutions are depending on non-national, non-sovereign organisations to preserve the national digital heritage.

Another interviewee voiced concerns about what content is actually being captured, let alone preserved: 'there is a huge amount of digital material that is not being captured

⁷ Una Walker, National College of Art and Design.

⁸ National Library of Ireland, 'Born digital', available at <http://www.nli.ie/en/born-digital.aspx> (accessed on 1 August 2012).

⁹ *Ibid.*

¹⁰ See <http://archive.org/web/web.php> (accessed on 1 August 2012).

by anyone', a problem exacerbated by the fact that 'correspondence...and interaction between people' have changed.¹¹ Many institutions expressed particular concerns about the preservation of e-mail correspondence: it was not clear how such correspondence is being preserved and how the ethical issues associated with preserving and accessing it will be resolved. One respondent felt that people used e-mail, erroneously, as a record-keeping system to preserve their business records. Yet, without the policy-based use of e-mail archiving solutions or other software, at an institutional level there are no guarantees that important correspondence encapsulating institutional memory will be captured or preserved.

A number of archives, including the Irish Architectural Archive (IAA), already receive born-digital material and are having to face the problem of archiving and preserving digital collections that are unstructured, undocumented and more complex than their paper-based counterparts. The IAA is taking a very pragmatic and proactive approach to this problem and is looking to adapt existing technology to build a digital repository that can handle the complex issues associated with file formats (this is further discussed in 'Digital file formats', below). This development, it hopes, will reduce the loss of important organisational information and help the archive to grow its collections, pre-empting the surge of born-digital content while resolving the cataloguing and ingestion issues of said content. Colum O'Riordan, Archive Administrator at the IAA, hopes that the archive's problem of ingesting and accessioning born-digital material could be largely resolved, once practitioners and architectural practices use the repository to store current projects and adapt their workflows to allow access and repurposing of content.

Although changing people's workflows now will help with future accessions of born-digital collections, archivists currently face dealing with important digital collections that do not follow any consistent data management plan or approach. One archivist commented that the accession of a collection of CDs, representing an artist's life work, into the archive was 'classic of everything that is problematic with digital preservation'. The CDs contained:

[a] random file structure [with] random file names, all shortened to some sort of self-created code which evolves...from disk to disk [and a] very complicated directory structure with probably somewhere between 50 and 60 per cent of the directory...empty. Now, are they supposed to be empty, did he forget to fill them? We don't know.

¹¹ Irish Traditional Music Archive.

More problematic for digital preservation was the fact that no metadata or documentation accompanied the collection, and therefore basic, yet essential, information was missing, such as what programs were used and what versions. Fundamentally, the 'meta-data was zero'.¹²

Another archive received the office contents of an organisation that has ceased to operate, bringing ethical as well as technical difficulties: 'we took in a large collection of [the organisation's] archives and basically...it is in a complete jumble'.¹³ The archive received 'the contents of the hard drive of the computers' but has yet to catalogue the material and is faced with a new challenge, that of archiving a born-digital collection. This raises the question: in accepting a collection or archive, what level of responsibility for managing the data does the recipient assume? Kasandra O'Connell, Head of the Irish Film Archive (Irish Film Institute), spoke of a recent experience with a famous director who approached the IFI with floppy disks from which he could not retrieve the files, not only because of the media type but also because of the format. The IFI was able to recover the content, and the owner was able to see scripts that had not been seen in years.

The DRI research team asked interviewees about their preservation process and whether they had any written procedures or policies for future-proofing their content. A number of stakeholders indicated that they have in place or are in the process of developing digital preservation strategies or policy documents, and although the majority have yet to formalise their procedures, they are acutely aware of the implications and problems of digital preservation. In a few cases our interviewees had not previously considered written policies on preservation practices but said that the question raised an important issue that they were now prompted to address. One respondent felt that the DRI had a strong role to play in this area:

When I see the word[s] Digital Repository Ireland, I would expect to find born-digital records are stored there and preserved there so that they can be migrated forward into new formats and then preserved and made accessible at the right time. And I really think that is where the gap is more than any other gap.¹⁴

While archivists are often struggling to resolve these issues, there was also a clear sense that there is a lack of awareness of good digital preservation practice among the general

¹² Anonymous institution.

¹³ Anonymous institution.

¹⁴ Anonymous institution.

public. The IFI launched a short public appeal on YouTube about its preservation of Irish film records. Many of the suggestions documented on the YouTube site indicated a lack of understanding of digital preservation processes: 'all they have to do is go down to PC World and buy a few hard drives and then they can throw all the film out'; put it on YouTube and 'just get rid of copyright, that will preserve [it]'. In response, Kasandra O'Connell spoke of the Irish Film Archive's role in education and in increasing public awareness of digital preservation. She asserted that people need to be aware of the digital content that they personally hold and be proactive about its accessibility in the long term. She cited the example of photographs: 'Do you know where they are? Have you called them proper file names? Are you looking at them every couple of years?'. The IFI's experience indicates that there is a clear need for education and training on digital preservation. The lack of awareness among the general public is worrying, as it indicates, as do some of the experiences cited above, that much content is currently being lost.

Storage requirements

Directly linked to the challenges of digital preservation is the issue of digital storage. The storage requirements of our interviewees were diverse and ranged in size from 4 gigabytes for one archive to 65 terabytes for another. Despite the diversity in storage needs, which is linked to the type of content held by individual institutions, respondents faced similar storage issues and problems. Many informed us that they have to review their current storage procedures because of increasing demands. One described a common problem: 'the amount of space and storage needed for...digital material...is causing havoc with...IT departments'.¹⁵ These storage demands are linked to digitisation activities, an increase in the amount of born-digital data, including research data and academic outputs, and developments in

¹⁵ Anonymous institution.

media, as well as the adoption of increasingly complex data types. Additionally, as Brian Rice, Archivist at the RTÉ (Raidió Teilifís Éireann) Sound Archive, stated: 'storage... has become cheaper; standards and expected standards have increased'. These expectations impact on how data can be accessed, used and repurposed, as archives are faced with the problem of moving and processing data files that are growing substantially. The increased demands on storage reflect international trends and a global increase in the amount of data being produced and processed. The scale of this issue is highlighted by the fact that less than a decade ago 'there were only 5 exabytes of data online' but in 2010 'estimates put monthly Internet data flow at around 21 exabytes'.¹⁶ Our interviewees, therefore, are faced with the problem of ever-increasing storage demands, coupled with increased user expectations and the long-term preservation of massive datasets and files.

One institution estimated that one of its collections, which contained archival-standard TIFFs, was 40–50 terabytes.

As highlighted above, 'standards and expected standards have increased'. A number of interviewees spoke of imaging to archival standards to produce high-quality, high-resolution images and the consequent storage problems. As one respondent informed us:

We image...to the best of our ability to conservation standard, which is [a] 21 megapixel image. But [this] creates a massive storage issue...We have the conundrum as to whether or not we compress the files...We have to assess each project from the outset, and up until now we have done a conservation-standard image. But now we are finding...we are having to go back and resize everything, which is time-consuming, to compress it all.¹⁷

One institution estimated that one of its collections, which contained archival-standard TIFFs, was 40–50 terabytes. Given that the RTÉ Sound Archive estimated its current 'total requirement...[as] somewhere in the region of 65 terabytes', this storage requirement for a single collection is notable.

¹⁶ Audrey Watters, *The age of exabytes: tools and approaches for managing big data* (2010), available at http://www.readwriteweb.com/archives/download_our_latest_free_report_the_age_of_exabytes.php (accessed on 21 August 2012).

¹⁷ Anonymous institution.

Although this comparison between one project and an entire archive may be an extreme example, many interviewees were facing difficulties associated with massive datasets. Anthony Corns, GIS/IT Manager at the Discovery Programme, described how archaeological models and surveys generate huge datasets, citing one aerial image that consisted of raw data containing 150 million high points. He explained that when these raw data are subsequently modelled and rendered, the model could be at least 50 gigabytes. An additional challenge is how to render and make these data accessible over the internet.

Overall, our interviewees had a wide range of storage requirements and challenges. This diversity is directly linked to the various types of digital, and analogue, material that our stakeholders are concerned with; for instance, the IFI, which has yet to go ‘down the route of digital preservation files’ using DCPs (Digital Cinema Packages), informed us that a two-hour film of preservation quality would be approximately 6 terabytes. Another participant, dealing with high-resolution TIFFs, lower-resolution JPEGs and associated text encodings, estimated that it held just under 4 gigabytes of content for two digital projects but added that a third project, which includes a significant amount of moving images, would on its own require 5 gigabytes. The Irish Traditional Music Archive, which predominately deals with audio files, estimated that its servers held 7 terabytes of data but that it would require significantly more storage by the end of the year.

Most of our interviewees had growing digital collections and were looking for new storage solutions and ways to cope with increasing demands. A number informed us that they were at the limit of their storage capacity and were investigating alternatives to on-site storage. The RTE Sound Archive spoke about its investigations into corporate storage but concluded that ‘the cost per terabyte and the pricing structure just made it impossible...to afford...[and other] options like cloud were just ridiculously expensive as well’.

Other interviewees also spoke of cloud storage, but only a handful actually used the cloud, for some, but not all, of their storage requirements. Events such as HEAnet’s launch of EduStorage in April 2012 and the release of the Data Protection Commissioner’s guidelines *Data protection ‘in the cloud’* in July 2012 propose the use of virtual storage as a viable solution. Yet, although some institutions were considering the cloud for their storage solutions, many expressed reservations. One asked if it was trustworthy, and another librarian expressed concern about the potential loss of control of sensitive user information if cloud-based services were in different legal jurisdictions:

The conditions under which the American security forces could access your record would be different to what they’d get access to here if that

issue is to arise. So, for example, we have Chinese students studying here. Let's say the government back home wanted to find out who they were...we would just say no. We can't say what the Americans or the Israelis would say.¹⁸

Despite these reservations, one respondent stated that 'removable hardware, LTO tapes [etc.]...are no longer recommended for digital preservation'; instead, 'cloud computing' is advocated.¹⁹ In terms of back-up and disaster recovery, however, many institutions informed us that they still used these methods, among others.






In conclusion, this review indicated four crucial policy challenges. Firstly, digitising projects need to be undertaken in a framework that ensures long-term preservation of the digital assets produced; this includes explicit workflows across the data lifecycle. Secondly, of great importance for the DRI are the concerns raised by institutions about the difficulties in preserving born-digital content, concerns that highlight a policy gap and the need for guidelines, in addition to technical solutions. Thirdly, clear guidance, in terms of shared guidelines, appropriate workflows and preservation processes, is necessary to ensure that today's digital objects will be accessible to future generations. Finally, although many institutions are currently able to meet their storage requirements, these requirements have increased exponentially and will continue to grow.

¹⁸ Anonymous institution.

¹⁹ Anonymous institution.

Digital file formats

Digital data can be stored in a variety of formats; for example, there are over 60 common file formats that can be used to store electronic text.²⁰ These formats vary in many ways, but from an archivist's perspective the key issue is whether the format will be accessible in the future. As high-quality formats are larger in size, another issue is how to maintain a balance between archival demands of high quality and storage cost limitations. The US National Archive gives the following suggestions to those considering which format to use:

-  The format is publicly and openly documented.
-  The format is non-proprietary.
-  The format is in widespread use.
-  The format is self-documenting.
-  The format can be opened, read, and accessed using readily-available tools.²¹

In recognition of how quickly formats change, the US National Archive does not mandate the use of particular formats. Other archives, such as the UK Data Archive, take a similar approach. Rather than delineating the only formats that it is willing to handle, it outlines which formats it prefers and which are acceptable.²² This pragmatic approach is in recognition of the fact that, at times, archives have little choice in what formats they are offered: if a highly important sound recording is deposited in a low-quality format, the archive has little option but to accept it. Some institutions, however, generate content themselves, and here decisions need to be made about which form content should be generated in. As we have seen in 'Digital preservation', above, many of the organisations interviewed

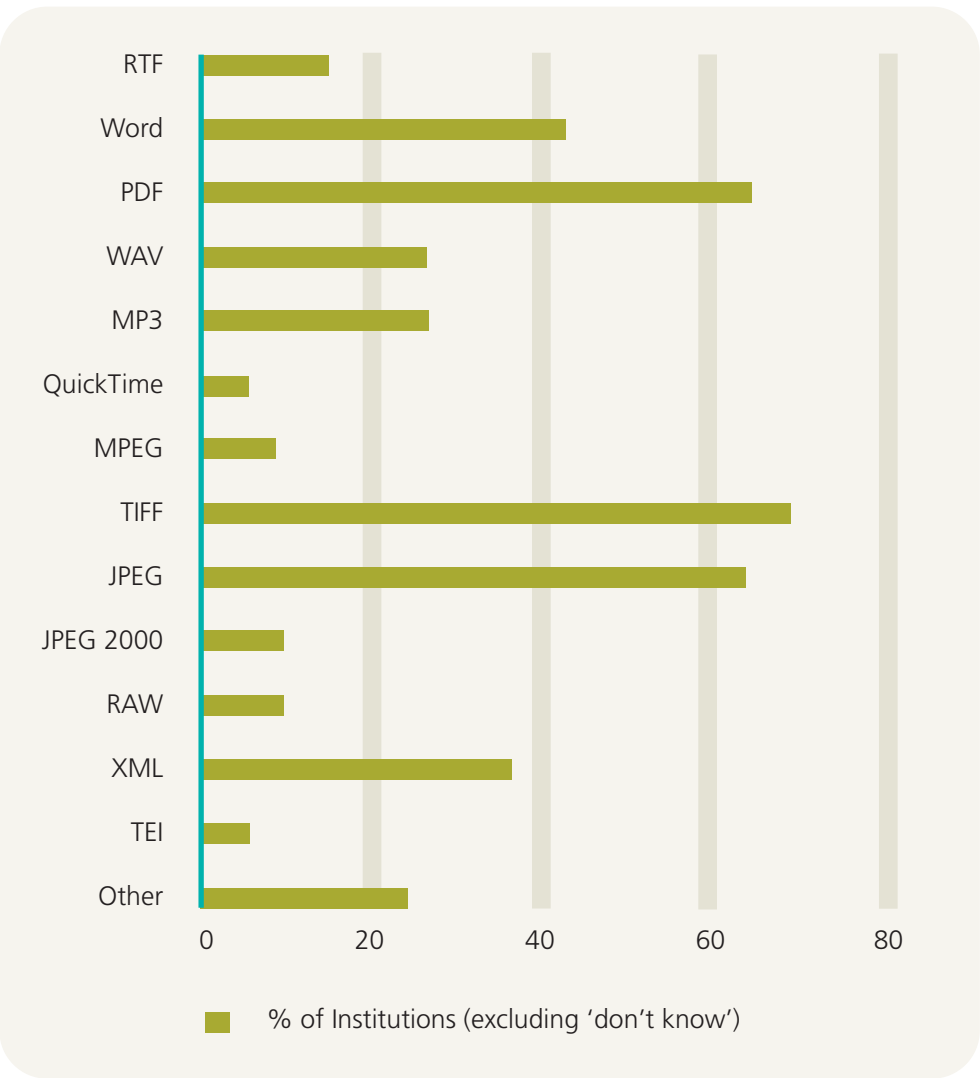
²⁰ Sources of further information on the formats mentioned in this section can be found in Appendix 2.

²¹ See <http://www.archives.gov/records-mgmt/initiatives/dav-faq.html> (accessed on 1 August 2012).

²² See <http://www.data-archive.ac.uk/create-manage/format/formats-table/> (accessed on 1 August 2012).

stored the same digital object in a variety of formats, a higher-quality format being used for preservation and a lower-quality format for sharing and access. Here, we outline which file formats the interviewees reported were held by their institutions.

Fig. 4: Formats used by institutions



From Fig. 4 it can be seen that the main format for textual data is PDF, with Microsoft Word in second place. Microsoft Word is a proprietary format, that is, it was created and is owned by a particular company (Microsoft) for use in its applications. The difficulty with proprietary formats is that should the company cease to trade, cease to support legacy software versions or change the nature of its software, it may be difficult or impossible in the future to access documents saved in that format: for example, documents written using the first word processor, WordStar, cannot be accessed on modern

versions of Microsoft Windows. As we saw earlier, one organisation still received documents in WordPerfect, a proprietary format that, since 2001, is available only for computers running Microsoft Windows (and OpenVMS). Many archives, such as the UK Data Archive, accept Microsoft Word documents, as they are so universally used, but are aware that this format may become obsolete over time.

Similarly, as PDF is so widely used, it is often accepted by archives; however, it is not considered an archival format. Although PDF was initially a proprietary format created by Adobe, it was released as an open standard in 2008. There are difficulties in using it as an archival format, however, because it does not embed the fonts that are used in the document. An archival version of the PDF format, PDF/A, was created in 2005 precisely to assist the long-term digital preservation of electronic documents. Microsoft Office and Open Office have introduced the ability to create PDF/A documents; however, they are not, as yet, widely used, and none of the institutions reported having PDF/A files. As with other archival formats, the files stored in PDF/A are much larger than those stored in PDF. In archival terms, RTF, while also a proprietary format, is preferable to Word and PDF, as it was created by Microsoft precisely to allow cross-platform sharing of documents and the specifications of the format have been published, allowing developers to include it in their software; therefore, it is far less likely to become unsupported in the future.

A number of the institutions that we interviewed held audio material, originating as broadcast material (RTÉ Sound Archive, RTÉ Raidió na Gaeltachta), field recordings of musicians (Irish Traditional Music Archive) or audio recordings of research interviews (Irish Qualitative Data Archive). Three audio formats were found in these collections: WAV, AIFF and MP3. WAV and AIFF are considered to be archival formats; however, the files that they produce are considerably larger than MP3 files (MP3 is not an archival format). WAV and BWAV are recommended for use in archiving by the International Association of Sound and Audiovisual Archives, while the UK Data Archive's preferred audio format is FLAC, but it considers either WAV or AIFF to be acceptable. If audio files have been donated in MP3 format, the archive may have no choice but to accept them. WAV files contain more information than MP3s, and it is therefore not meaningful to attempt to translate an MP3 file to WAV format. Indeed, one archive noted that when this was done (in error), an unwanted hum was added to the final files. AIFF, an audio format developed by Apple, was accepted by one of our interviewed organisations.

A number of the institutions held multimedia audio and visual formats. QuickTime files were held by NUI Galway Library and the IFI. QuickTime is a proprietary format created by Apple. MPEG-2, MPEG-3 and MPEG-4 formats were also used. MPEG is a standard

created by a working group of the International Standards Organisation and the International Electrotechnical Commission, and hence in archiving terms it is preferable to QuickTime. As with the audio formats above, however, institutions were not necessarily in a position to dictate which formats they would accept.

TIFF, JPEG, JPEG 2000 and RAW are all image formats. TIFF was the most popular format, used by 71% of the organisations interviewed. Files were imaged at 600 dpi, 400 dpi or 300 dpi. TIFF is considered to be the standard archiving format; however, TIFF files are considerably larger than JPEG files. As we have seen in 'Digital preservation' above, many archives produced JPEG versions for dissemination (particularly for use on the web) rather than preservation. JPEG 2000 was introduced in 2000 as an archival version of JPEG. It is not supported by many web browsers, however, and has failed to be widely adopted to date, although one of the institutions interviewed stored JPEG 2000 files.

A variety of organisations, including the Royal Irish Academy, the National Folklore Collection and NUI Maynooth Library, held RAW files. RAW files are the original image files created by digital cameras. They are a proprietary format, with each camera manufacturer creating its own version of RAW files. DGN (Digital Negative Format) is an open format created by Adobe with the aim of becoming a standard format into which RAW files could be converted for archival use. It is based on the TIFF format, although it has not yet been adopted as a standard. None of the institutions reported using DGN, and they may find it difficult in the future to access the RAW files that they hold.

Thirty-eight per cent of the organisations reported holding XML (Extensible Markup Language). XML was developed by W3C (World Wide Web Consortium) and is a non-proprietary, platform- and software-independent mark-up language used for data and document encoding, storage and transport. TEI (Text Encoding Initiative) is a set of XML-based guidelines for the digital encoding of literary and linguistic texts.

A number of organisations store formats that were special to their communities. The IFI holds DCP (Digital Cinema Package) data. DCP is a format created by a coalition of the major film studios to collate audio, image and data streams within a single format. The IAA holds CAD and BIM (Building Information Modelling) files, which document a range of digital information about the building process, from plans to manufacturers' details of a building component. The Discovery Programme, which creates 3D imaging of archaeological sites, stores 3D PDFs and raster data. It creates images using WMS (Web Map Service), a widely used format for transmitting geo-referenced map images over the internet that was created in 1999 by the Open Geospatial Consortium. Additionally, it uses WFS (Web Feature Service) and WCS (Web Coverage Service), which

are specifically designed to assist in the delivery of digital geospatial content over the web. Clare County Library has considerable map-based resources on its website.²³ These are stored as SVG files, an XML-based format for 2D vector graphics. Additionally, it stores maps in DjVu files, which have been created with DjVu image compression software. DjVu is an open file format and is used by the Internet Archive for its Million Book project. For those organisations that manage CAD or 3D modelling files, such as the IAA and the Discovery Programme, there are as yet no archival formats for these files. This problem is compounded by the use of proprietary software and further complicated by different software and application versions and releases. Colum O’Riordan, Archive Administrator at the IAA, spoke of the difficulty of dealing with CAD files, not only because there is no archival CAD format but also because backward compatibility is poorly supported. More challenging, perhaps, is the fact that the same CAD file opened in Microsoft Windows, Linux or Apple OS will not necessarily look the same. This is a serious problem for long-term preservation, especially because architects are now designing in 3D, moving away from CAD and into BIM, a move that will see architectural archivists and repositories dealing with the legacy of different design systems.

In conclusion, although many file formats exist, a relatively small number were in use in Ireland. There were commonalities in terms of the formats used for textual, image and audio data. The formats used for dealing with geospatial and broadcast material were not widely found across the institutions reviewed. This was a reflection of the uniqueness of this content in the Irish context; the formats used for these data were in common use internationally. Some archives (such as those dealing with 3D data or CAD files) faced a particular problem as preservation formats were not available. Although cost pressures could drive institutions to select non-preservation formats, in the main this was not evident. Instead, two types of format were often held: high-quality, large-size preservation formats and lower-quality, smaller-size formats for web dissemination. Those institutions whose remit was not primarily in the field of archiving (such as art galleries) sought guidelines on which formats would be most appropriate to preserve the data in their collections. A national preservation policy would also have to provide guidance on organisational policy and workflows to track the development of new formats and migration of data to new formats where necessary.

²³ See <http://www.clarelibrary.ie> (accessed on 2 August 2012).



Metadata and vocabularies: describing data

Metadata

Metadata are often described as 'data about data'.²⁴ The term refers to information attached to an object that tells us more about the object. A library catalogue record about a book is a form of metadata. Metadata can be documented in many different ways; metadata schema or standards are common or shared ways of documenting metadata. These schemas have emerged in different domains, as the key characteristics of the objects described vary according to the communities using the objects.²⁵ Shawn Day, Project Manager at the Digital Humanities Observatory, outlined the problem faced by many:

one of the biggest barriers we obviously face is just standards and formats. You are really dealing with the multitude of the ones out there. Part of this would have to do with...the fact that we are dealing with projects that have...been creating stuff over the past 15 years. Both within a world where these things are changing anyway but also in a situation where there has been no central authority by any stretch that people could consult to determine what are best practices.

In the digital world, metadata are important because if online collections use the same metadata schemas, it becomes possible to search across collections.

One-fifth of those interviewed were not aware of which schema they were using. The results in Fig. 5 are from those who reported their metadata use, and most institutions were using between one and four standards. Additionally, 6% of the institutions reported that they did not use any international standard but instead developed an in-house

In the digital world, metadata are important because if online collections use the same metadata schemas, it becomes possible to search across collections.

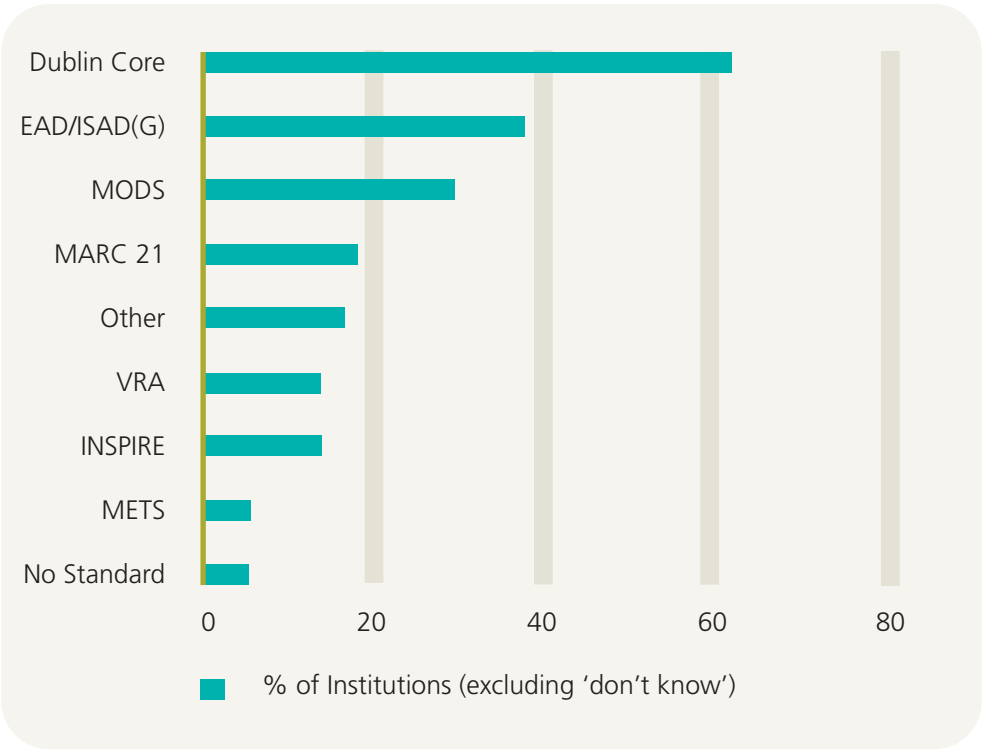
²⁴ Sources of further information on the metadata standards and vocabularies mentioned in this section can be found in Appendix 2.

²⁵ A metadata record, therefore, typically consists of a number of predefined elements representing specific attributes of an object, and each element can have one or more values; for example, most metadata schemas will have an element called <Title>, which refers to the name of the object that is being described.

description framework or were currently considering which standard to use. It is important to note that, once modified, 'standards' are no longer 'standards', which limits interoperability and thus the potential to share metadata between systems and organisations.

Dublin Core is one of the simplest metadata schemas ('Dublin' refers to Dublin, Ohio, which was the location of a workshop in 1995 from which the schema was developed). It consists of fifteen metadata elements, including title, creator, subject and description. As it is relatively versatile, it is widely used internationally. Indeed, it was the most popular metadata schema used by the interviewees, with 61% reporting that they used it. The next most popular metadata schema was ISAD(G), which was used by 39%, predominantly libraries, reflecting ISAD(G)'s (EAD) origins. This schema was developed for use in archives by the International Council of Archives, which published the first version in 1993. This schema contains 26 elements, of which six are mandatory: reference code, title, name of creator, dates of creation, extent of the unit of description and level of description.

Fig. 5: Metadata standards



The MARC format, used by 19% of the institutions interviewed, was created by the US Library of Congress in the early 1970s for use by libraries. The MODS standard is another initiative of the US Library of Congress and was established for use by libraries in describing bibliographic items. It was created as an alternative to the complexity of the MARC

format and the simplicity of Dublin Core. The MODS schema was used by 29% of the institutions. The METS schema, which was used in 6% of cases, also emerged from the US Library of Congress.

Metadata are field- or domain-driven, and therefore institutions that hold or use geospatial information are increasingly adopting INSPIRE, a standard for spatial information that was developed by the European Union in 2008. It was used by 13% of institutions, including the All-Island Research Observatory (AIRO), which provides map-based interfaces for government datasets, and the Discovery Programme, which curates 3D maps of archaeological sites.

Another standard that featured in some of the interviews was VRA, which was developed by the Visual Resources Association in 1996 to aid the description of visual objects. It was used by 13% of our interviewees, especially those, such as the National Gallery of Ireland, whose collections were made up largely of images.

The other standards in use (by a single institution in each case) were ESE, NISO MIX, IPTC, IEEE LOM, EBU Core and SPECTRUM. Two institutions mentioned adopting MODS in the future, and one mentioned future use of DDI (a standard applied to social science data). As we saw earlier, many collections contained digital data in a wide variety of languages. In most instances the metadata reflected the language of the original object (so, for example, an Irish-language song would have metadata in Irish).

Controlled vocabularies

Controlled vocabularies and thesauri are used in archiving to ensure that objects are described in common ways: e.g., a ‘worker’ could also be called an ‘employee’; the ‘labour market’ could also be described as the ‘job market’. For this reason, archivists have developed controlled vocabularies, thesauri and ontologies to give guidance to those adding data about an object (an ontology is a specification both of terms and of the relations between them).²⁶ Finding multiple instances of the same object becomes very difficult if that object is referred to in different ways on different records. For Irish data, a particular problem was noted in that multiple spellings were commonly used for many Irish surnames and place names (especially townlands). Damien Gallagher, at the time of the interview a software developer with An Foras Feasa, NUI Maynooth, faced this problem when importing an international database:

²⁶ For example, in the Getty Art & Architecture Thesaurus, ‘churches’ are a subcategory of ‘religious structures’, which are a subcategory of ‘single built works by function’.

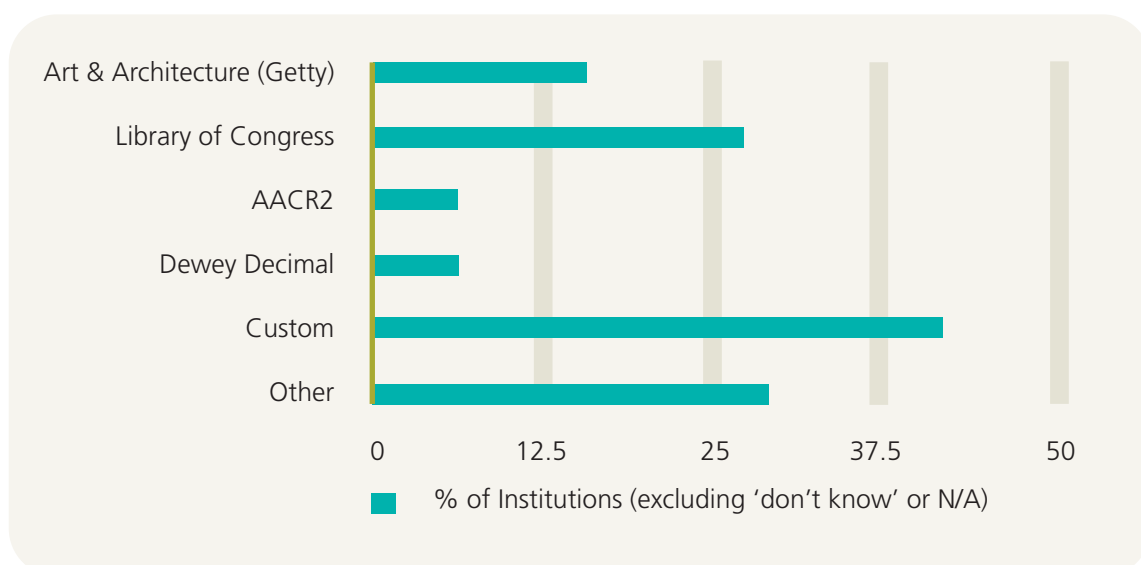
A lot of the records weren't written by the Irish people; they would have been written by French people or Spanish people, and they would have completely...misspelled [Irish names]. They would 'Frenchify' a certain thing. For example, I have seen Sweeney spelt in the French record as Suiney.

Críostóir Mac Cárthaigh, Archivist at the National Folklore Collection at University College Dublin, referring to a pilot project to index electronically the names, surnames and addresses of contributors to the National Folklore Collection, outlined the problem:

If you were searching for O'Sullivan, it would give you Ó Súilleabháin and the different various spellings of Súilleabháin. It was a big challenge as well because, in Ireland, no two people spell the townland the same.

Figure 6 below shows the considerable variation in the guides used, with 33% of institutions creating their own. The 'Other' category includes the following controlled vocabularies and guidelines: the UK Data Archive's HASSET Thesaurus, the Placenames Database of Ireland (www.logainm.ie), the Irish Public Service Thesaurus, the International Federation of Film Archives Taxonomies, the UK Archival Thesaurus, the Dictionary of Irish Biography and the British Architectural Library's Architectural Keywords. It should be noted that although some of these vocabularies are freely available, others are proprietary and require licensing for use, which can be expensive.

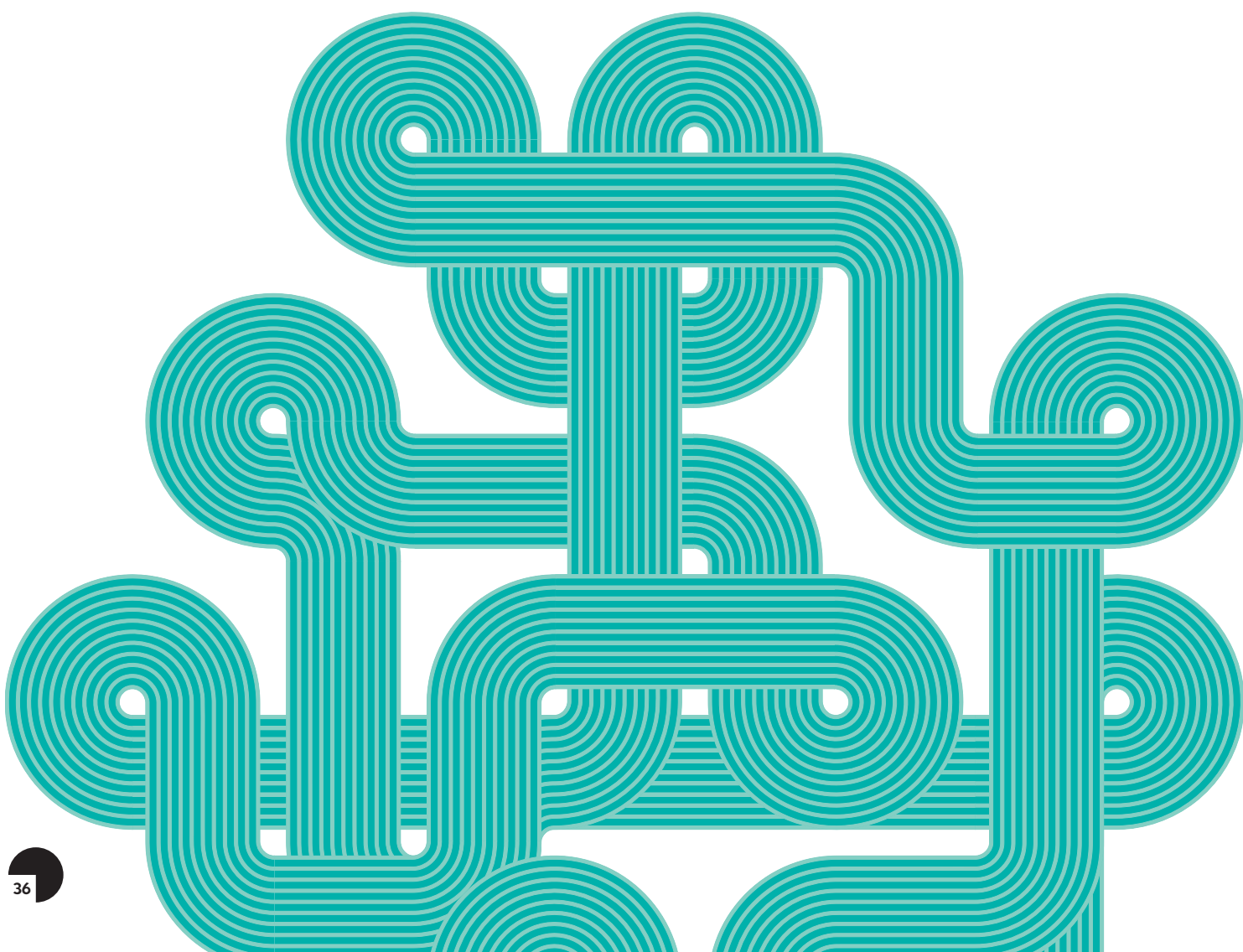
Fig. 6: Ontologies, thesauri and fixed-word vocabularies used



This review found that the majority of the institutions were using at least one of seven standards and that Dublin Core was by far the most popular. These standards were appropriate to the content and were also used by special user communities internationally; for example, libraries and archives

both in Ireland and internationally tend to use ISAD(G). A number of the institutions, however, used standards that were domain-special, e.g., INSPIRE, EBU Core and LOM, reflecting the uniqueness of the content that they held. This highlights the challenge faced by content holders in identifying which standards are most appropriate to their field: the temptation to develop one's own bespoke approach should be avoided, as it may limit future data connectivity. The challenge facing the DRI will be to facilitate interoperability between datasets developed in different domains. It is highly unlikely that the DRI will be able to support all of the metadata standards in current usage; however, it plans to support a suite of standards that balance best practice and common usage with current national investments.

The temptation to develop one's own bespoke approach should be avoided, as it may limit future data connectivity.



User tools

Collecting, managing and preserving digital material are crucial activities in the lifecycle of born-digital and digitised material. In order to engage users and to add value to these collections, however, many of our interviewees created and provided user tools to give audiences new ways to interact with digital material. These tools include user-generated annotations, crowd-sourced transcription correcting, networked mapping and graphing of relationships between material, text comparisons, data and geospatial visualisations, interactive maps, online exhibitions, interactive guides, interactive tables and educational tools. Mobile applications ('apps') were also developed, ranging from augmented reality to online catalogues and other visualisation aids for resource discovery, including timelines to narrow search criteria and results. Our interviewees were cognisant of the need to provide user tools, not only to provide new ways to access digital material but also to allow content to be reused, repurposed and reinterpreted, adding value to the content, as well as the archive. The development of user tools is in response to a noticeable shift in user expectations. Of this, one respondent stated that just using:

The development of user tools is in response to a noticeable shift in user expectations.

A PDF format...is not good enough anymore in terms of how people interact with [material]...[New] undergraduate[s] that come in...access and use the material in a very different form than perhaps what we created even three years ago. And how they expect it to be delivered, how they want to manipulate it, how they want to use it [have changed].²⁷

Providing new opportunities for users to engage with content also poses new challenges to content holders and tool developers, as they must grapple with the problem of providing sustainable user tools. This maintenance of functionality is another major issue for long-term digital preservation. Content holders must not only preserve digital objects but also provide long-term access to the context in which the objects are viewed,

²⁷ Anonymous institution.

visualised, used and manipulated. Copyright restrictions can also constrain what end-users can do with digital material, while funding and resource allocation to develop new user tools and resources is a major issue faced by all.

Exploring digital content: user tools for content engagement

The majority of our interviewees were identifying new ways to increase user engagement with the data in their collections. While access to material, both digital and analogue, is primarily achieved through the development of finding aids, 60% of our interviewees provided additional user tools to support resource discovery and enhance the user's experience with digital content and material. Although the remaining 40% did not yet provide any additional user tools, the majority were considering future developments in this area and a number discussed working with external partners.²⁸ Arlene Healy, Manager of Digital Systems and Services at Trinity College Dublin Library, identified user tools as an area for future development, stating that 'right now, our digital library displays the images after being browsed or searched so there is no...value added' to the objects, but she continued that 'it is definitely a future requirement'.

The integration of user tools into our interviewees' websites and systems include: in-house or bespoke development of applications; the use and customisation of open-source tools; tools developed by or outsourced to a third party; and proprietary, off-the-shelf solutions. Of the user tools, 50% were developed in-house, 25% were based on open-source solutions and 25% were sourced from proprietary software companies. In 28% of cases, developers used a combination of open-source, proprietary and bespoke tools.

Proprietary tools used by institutions included Zoomify, Tableau, InstantAtlas, 3D Issue, as well as a Flash-based exhibitions outsourced to a third party and tools or modules provided through systems such as eMuseumPlus. Open-source tools included OpenZoom, OpenStreetMap, TimeMap, Sibelius Scorch, Fedora GSearch, Solr and the use of Google Maps to layer historical maps and images. Tools developed in-house included discussion forums, user annotations, geo-mapping and spatial analysis.²⁹ The Irish Qualitative Data Archive's anonymisation tool was developed in-house and is free to use, representing an important development in assisting researchers in preparing social science data for archive and reuse while meeting strict ethical standards.

²⁸ Use of social media is not included in these figures.

²⁹ For example, see AIRO's census mapping tools, available at <http://airomaps.nuim.ie/flexviewer/?config=Census2011.xml> (accessed on 31 August 2012).

Curated exhibitions: contextualising and interpreting material

The tools mentioned above are an important way for users to engage with content and can enhance resource discovery, as well as the user's experience. Another method to achieve this, and to add value to a collection, was through the development of online curated exhibitions of digital assets, which provided users with novel ways of interacting with primary material. This method also provided the host institution with an opportunity to display collections that might otherwise be temporary floor exhibitions or might not be displayed at all. The National Archives of Ireland, the National Library of Ireland and RTÉ were among a growing number of institutions to provide this feature to their users, while a number of interviewees, including the IFI, shared their plans to develop this feature in the future. Malachy Moran from the RTÉ Sound Archive told us that online curated exhibitions are a 'way of giving value [to] the archive without compromising responsibilities to preserve the material' while providing an entertainment element to the archives. He added that designing an online exhibition required considerable resources and a great working knowledge of the collections but felt that the return was worth it, especially as online exhibitions helped to fulfil the archive's remit of providing public access to its holdings.

Curated collections or exhibitions enable content managers to expose the richness of their collections while maintaining control over what content is actually made available. This is a useful method to employ, especially in situations where copyright or access policies, issues and restrictions persist. They act as an advertisement for an institution's holdings and can result in increased footfall, as online exhibitions, which often link multiple types of media objects, reveal the diversity of the content holder's collections. As an example, a podcast of a lecture linked to supporting documents, audio snippets or moving images showcases the archive but also enhances the users'/students' experience and their engagement with knowledge. An Foras Feasa at NUI Maynooth has already successfully developed and enhanced its repository with such features. Damien Gallagher, currently a senior software engineer at the DRI but previously the senior developer for An Foras Feasa's CRADLE (Collaborate, Research, Archive, Discover, Learn, Engage), explained the use of 'RDF [Resource Description Framework] to describe the relationships between each object in [the repository]', in order to generate graphs to illustrate the connections between different types of objects, including documents, audio, moving images,

**Curated collections
or exhibitions
enable content
managers to
expose the richness
of their collections
while maintaining
control over what
content is actually
made available.**

user-generated annotations, digital articles, PDFs and slides. Other tools developed in the current Flash-based system (a move to HTML5 is being considered) include a discussion forum and a slideshow tool. Similarly, the National Library of Ireland produced a Flash-based online exhibition for the 90th anniversary of the 1916 Rising and stated that this was an area that it would like to develop. It viewed online exhibitions and curated collections as an important platform for contextualising and interpreting its material. The Irish Museum of Modern Art has also produced a number of virtual tours that document past exhibitions and collections.

Online curated collections or exhibitions can also support cooperation between different institutions, and one respondent identified this as an activity that the DRI should support.

Going mobile: creating 'mobile-friendly services'

The development of mobile applications was discussed by a number of institutions. Although few had fully developed mobile applications that are available to download and use, many referred to this as a feature for future development, to complement and enhance their digital, as well as physical, assets and holdings. The University of Limerick informed us of its 'Ireland under siege' app, released in May 2012 and developed in conjunction with NUI Galway and the Royal Irish Academy.³⁰ The app, which brings to life the landscape of important battle sites, was developed for Apple iOS as well as Android, and is supported by a website and a blog, 'Immersive learning in history', which documents the 'production of [the] augmented reality mobile phone application'.³¹ The mobile app and the website encourage 'enquiry-based learning in historical research',³² while the blog, which reveals some of the technical choices and features, encourages users and students to engage with software development.

Libraries also described the development of 'mobile-friendly services' and apps to provide access to catalogues, as well as other resources. One librarian informed us that 15–20% of the catalogue's hits were now from mobile devices, and therefore mobile apps and QR codes represented an important way to keep up with user demands and the changes in user activity. Clare County Library also discussed developing 'a version of [its] catalogue for mobile devices' and was considering developing a mobile app for its map collection.

³⁰ See <http://www.irelandundersiege.com> (accessed on 21 August 2012).

³¹ See <http://testarea.edublogs.org/about/> (accessed on 21 August 2012).

³² 'UL launches "Ireland under siege" mobile phone app', available at <http://www.ul.ie/news-centre/news/ul-launches-ireland-under-siege-mobile-phone-app> (accessed on 21 August 2012).

Visualising geospatial data and other mapping tools

The use, creation and visualisation of geospatial data, as well as geo-browsing, were discussed by 16% of interviewees, including Justin Gleeson, Project Manager of AIRO, which gathers and analyses spatial datasets from across Ireland. He informed us of AIRO's developments, which included the use of different mapping and visualisation tools, and stated that it had considered various open-source tools to help with the visualisation and interactivity of datasets but had opted for proprietary software, InstantAtlas and Tableau, as a more cost-effective solution because it provided user support. AIRO, however, also used open-source software such as Drupal and MySQL and developed a Flash-based tool to query datasets. Examples of AIRO's mapping tools include Census Mapping, a Crime Mapping Toolkit and a Social Housing interactive map and analysis tool.³³

The Clare County Library website includes a tool, Clare GMaps, which uses Google Maps to layer historical maps of Clare over current Google images. The site also includes a tool called MapBrowser, which provides users with access to a number of historical maps, including David Rumsey's maps of Clare. The Oral History Network of Ireland also referred to a recent oral history project, which plotted oral accounts of a particular area to a specific location.

Building an online community: social media

Nearly all of the institutions we interviewed used social media to engage and interact with the wider public and their online user base. Social media was used to notify audiences of events, publications and other news, but it also provided institutions with an opportunity to encourage and develop relationships between and among members of their online community.

In a few cases, institutions that used social media often had dedicated staff to manage, populate and control the various sites. The National Library of Ireland informed us that its use of social media sites such as Facebook, Twitter and Flickr was of huge benefit. It not only attracted readers to the library but also helped it to enrich the catalogue's metadata: for example, users of Flickr left comments and in some cases identified unknown places, landmarks and individuals in historical photographs. Users also helped to date certain images and tagged and identified material objects, such as pioneer pins and flags. Without this type of user-generated content, much of the context of these historical photographs would never have been retrieved. This type of user engagement not only enriched the National Library of Ireland's photographic

³³ See <http://www.airo.ie> (accessed on 21 August 2012).

collection but also provided users with a unique opportunity to engage with the library's content and to contribute to and enhance their national cultural (digital) heritage. The National Library of Ireland also informed us that, after the publication of an article in the *Irish Times*, its Flickr Commons' photostream received 40,000 hits in one day alone, demonstrating how social media can facilitate greater user engagement with archival material (in contrast, in 2010, an average of 507 people per day visited the library's premises on Kildare Street).³⁴

Future developments of user tools: fulfilling user expectations

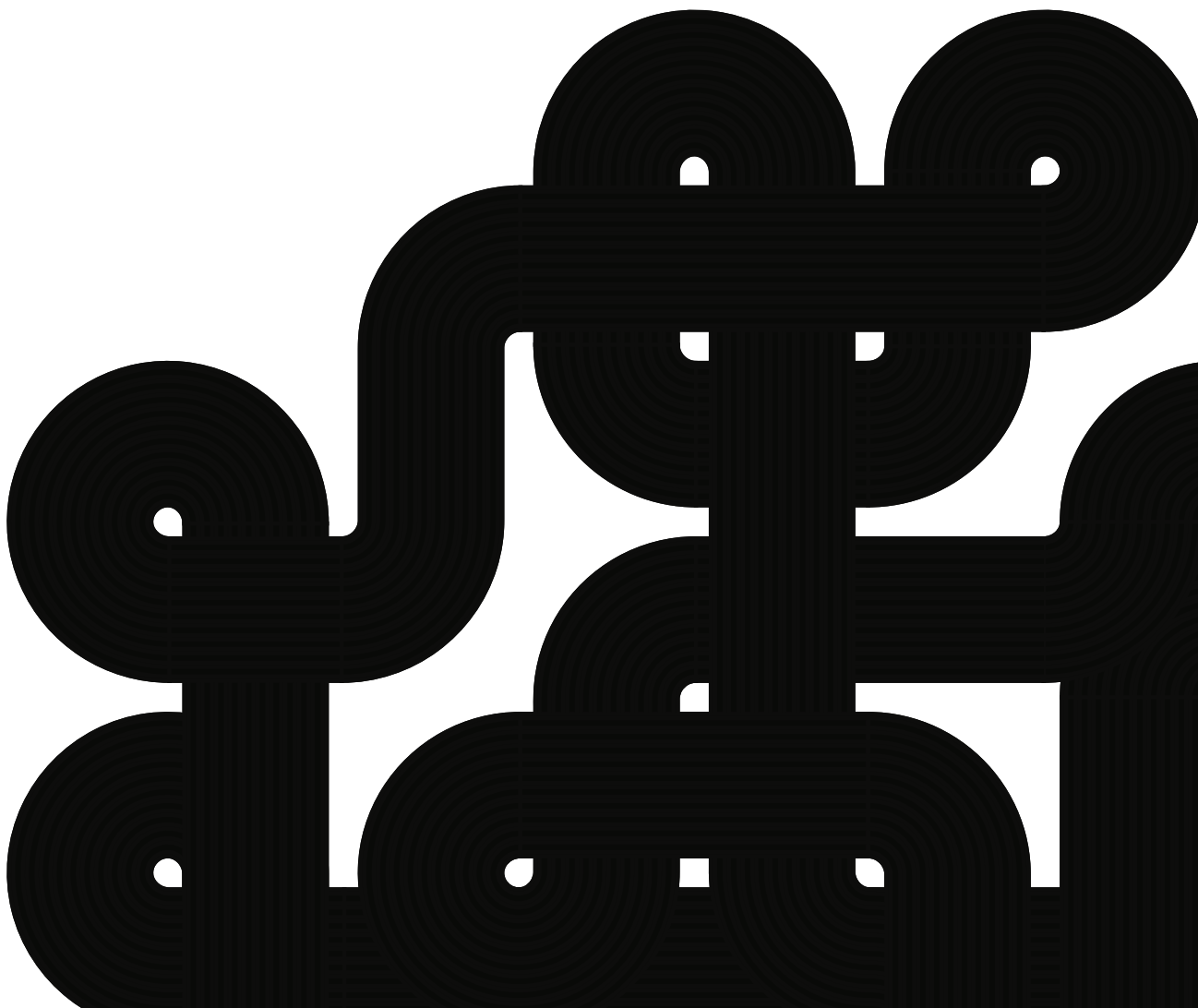
New user tools and resources to engage the public and the younger generation, including primary, secondary and third-level students, as well as academics and scholars, with digital content and enhance user experiences were an area that our interviewees identified as crucial for future development. Particularly, a number of organisations sought to build educational tools including games and other interactive resources. The provision and development of new learning objects and other educational resources was also mentioned. Temporal and spatial tools were also discussed, including the development of time maps and timelines to compare documents, audio etc. over different time periods, as well as geographical mapping of content to map collections and the landscape to allow users to interact with content through a map interface. Mapping tools were also discussed in conjunction with the use of crowd-sourcing to develop collections of special content (e.g., images of special architecture). The development of user workspaces, or 'light boxes', that allow individuals to curate their own private collections or exhibitions was also discussed by some of our interviewees, especially those working with vast photographic or art collections. Resource discoverability was also high on organisations' agendas, and a number of librarians spoke of future developments to optimise search queries and results and to support distance reading. A number of our interviewees wished to develop mobile applications, including apps for smartphones and tablets. One respondent noted that, although this was an area that it wanted to develop, it would look for platform-neutral applications to counter socio-economic divisions. Although our interviewees did not discuss the use of third-party APIs (application programming interfaces) for the creation of user tools, this is an important area of resource development. APIs also enable different systems to share metadata and data as they support the harvesting of content from trusted partners and institutions.

³⁴ 'Irish museums and galleries boast 3 million visitors in 2010', *Journal.ie*, February 2012, [http:// www.thejournal.ie/irish-museums-and-galleries-boast-3-million-visitors-in-2010-90608-Feb2011](http://www.thejournal.ie/irish-museums-and-galleries-boast-3-million-visitors-in-2010-90608-Feb2011) (accessed on 31 August 2012).

In this review, there was a strong sense that the development of user tools was a priority, and as discussed, many interviewees have already embarked on providing their users with innovative and novel ways to access, use and visualise digital content. Others wished to develop and enhance their digital collections with user tools and resources but were unsure of how they should do this. Discoverability and access to resources and objects was identified as

of key importance, and the use and development of different user tools with various functionalities was viewed as one method of achieving this goal. As an 'interactive' digital repository, the DRI will provide a number of specific, targeted user tools and will also support the development of resources through its API. This activity will enhance and enable collaboration between the DRI and the community while enriching the end-user's experience, thus achieving a goal shared by our interviewees.

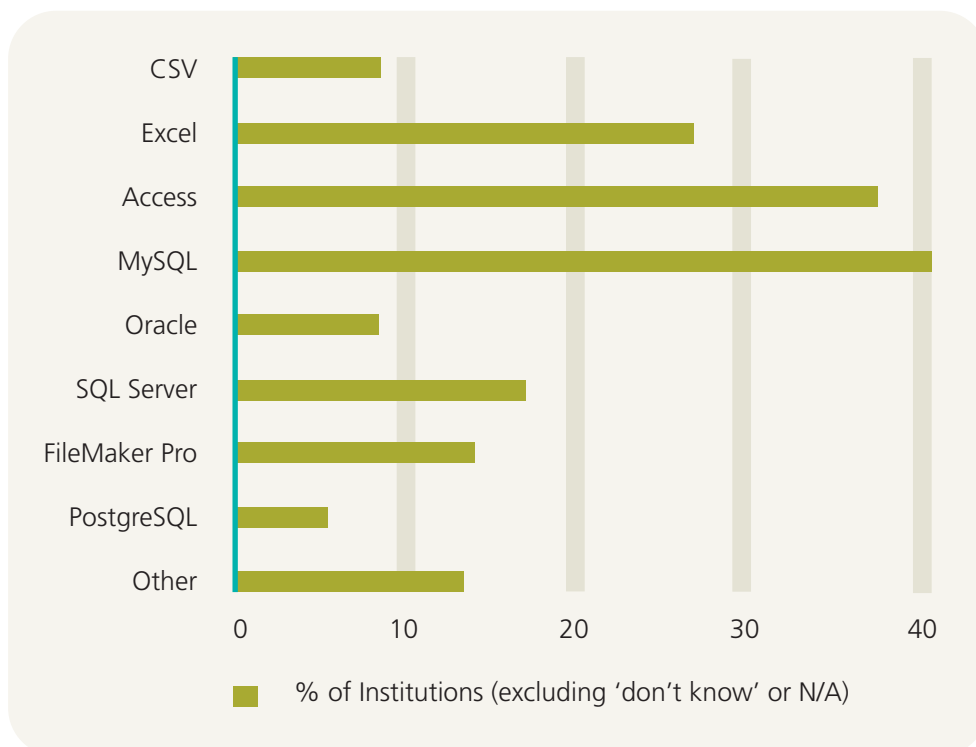
**Discoverability
and access to
resources and
objects was
identified as of
key importance...**



Structuring content: database formats/systems, content management systems and repository software

Figure 7 provides an overview of the database formats and systems used by our interviewees.³⁵ Microsoft Excel and CSV files are deliberately included; although neither is a database system, they are used by institutions for a variety of reasons, e.g., Excel spreadsheets for in-house management, crowd-sourced transcriptions or legacy catalogues, and CSV for storing raw data such as GIS or supporting the import/export of data. The potential use of CSV may be higher, given that it is a common export format. 'Other' database systems include Mulgara, eXist-db, Basis and unspecified GIS database formats.

Fig. 7: Database formats and systems



³⁵ Sources of further information on the database and content management systems mentioned in this section can be found in Appendix 2.

Many of the institutions used more than one database system. Microsoft Access was used by 37% of our interviewees. In most cases, however, Access was a legacy database that was in the process of being upgraded or migrated to a new system. The IAA informed us that Microsoft Access worked well for its in-house catalogue, but its 'biggest bugbear' was that the catalogue 'is not available online' and 'Access is simply not robust enough to put online'. To resolve this problem, the archive is migrating to Adlib, which will enable online access to the catalogue. Adlib, a library management system, is also used by the National Museum of Ireland and the National Archives of Ireland. Another archive also used Microsoft Access for its in-house catalogue, which at present is available only locally in the reading room, but expressed similar aspirations to make the catalogue more widely available.³⁶ This archive's priority in choosing a new system was 'openness...that it would have open database connectivity', that is, it would have the ability to support interoperability and connectivity between systems.³⁷ Some archives had discussed the use of open protocols, particularly OAI-PMH, as a method to share and harvest metadata. The archive was also keen to understand what library management systems other institutions were using or migrating to, stating that it is important to have a 'broad national perspective on things...so if there are a lot of institutions moving...[in the same direction], we would move in a very coherent way'.³⁸

The prevalence of open-source software and solutions is also reflected in Fig. 7, as MySQL is the most common relational database management system (RDMS) in use at the moment. Open-source database solutions not only are affordable but also can have advantages over proprietary software in terms of long-term preservation of and access to database content. The use of RDMSs also reflects international trends and SQL's dominance in the last number of decades. None of our interviewees mentioned NoSQL ('Not only SQL'), however, which is designed to be web-scalable.³⁹

Open-source database solutions not only are affordable but also can have advantages over proprietary software in terms of long-term preservation of and access to database content.

³⁶ Anonymous institution.

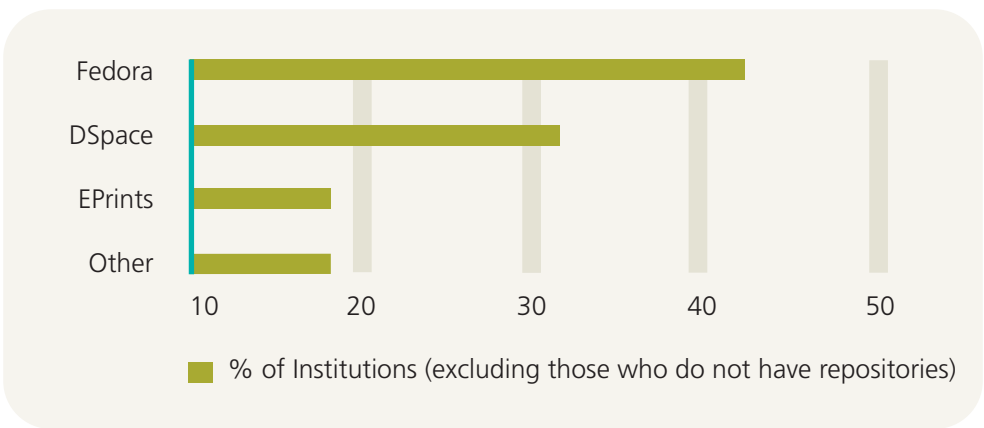
³⁷ *Ibid.*

³⁸ *Ibid.*

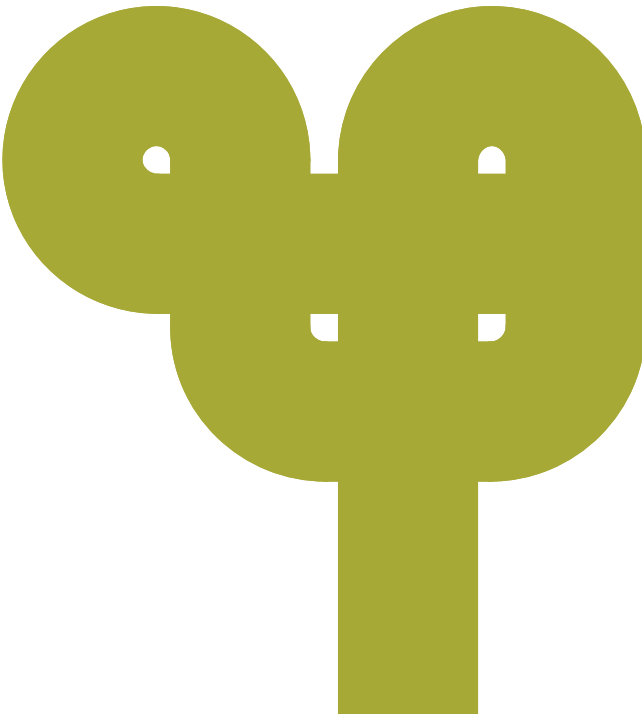
³⁹ Audrey Watters, *The age of exabytes: tools and approaches for managing big data* (2010), available at http://www.readwriteweb.com/archives/download_our_latest_free_report_the_age_of_exabytes.php (accessed on 21 August 2012).

Of the institutions that we have spoken to thus far, 38% had their own digital repository. Included in this number are institutional repositories that deal with academic output, e.g., theses and articles, but repositories also included other datasets, such as those for research-based projects or national cultural institutions. As Fig. 8 shows, the open-source digital asset management system Fedora Commons was the most popular, used by 44%. DSpace was used by 31%, and EPrints by just under 19%, both also open-source. Other repository systems included TYPO3, VTLs VITAL (a Fedora-based system) and other, unspecified commercial solutions.

Fig. 8: Digital asset management systems



Our interviewees used an array of content management and front-end systems, including commercial, open-source and custom in-house solutions such as Drupal, WordPress, Fez, ExpressionEngine, eMuseumPlus, ArtBase, Adlib and Kentico. It is noteworthy that there is no consensus on the content management systems used, reflecting the wide diversity and variety available.



Absences

There were some absences in the interviews (although it should be noted that an absence in the interview does not necessarily reflect an absence in practice). The interviewees did not discuss the preservation of data obtained or modified through the provision of user tools, nor policies that would ensure preservation and future maintenance of those tools. Tools were considered primarily as user tools, and not to facilitate ingest. There was little mention of possible intermediary machine formats or middleware that might be required by a front-facing application or of a desire for standard APIs.

There was no mention of the processes by which policies were developed and implemented, and few⁴⁰ mentioned delivering in-house education or training programmes aimed at raising staff skill levels. Linked data, an emerging practice to enable the web to connect related data that were not previously linked, was rarely mentioned.⁴¹ It is difficult to draw conclusions from these absences; although they could reflect absences in the institutions, they could also reflect a methodological limitation (put simply, not all of the questions that we asked could be answered by the person whom we interviewed). The absences are noted here, as the DRI will need to explore these issues in more depth in its future work.

⁴⁰ The Digital Humanities Observatory was a notable exception.

⁴¹ An exception is RTÉ, which is engaged in a linked data project proposal with the DRI and the Digital Enterprise Research Institute. In stakeholder consultation, Marie Wallace, Social Analytics Strategist at IBM, emphasised the need to consider a linked approach to sharing data.

Conclusion

A variety of institutions in Ireland are tasked with caring for digital content. This content is varied in nature; much is rare, unique and valuable. In the face of a rapidly changing technological field and considerable resource limitations, organisations are cognisant of the necessity to develop robust workflows in order to protect their digital collections and ensure that their richness is available to future generations. Many organisations are eager to add value to the resources in their care. Digital archives are moving beyond providing simple 'access' towards identifying ways in which they might transform their content to meet changing user needs.

Many are aware of the challenges that they face. Digitisation and digital projects are often based on short-term funding, with few resources available to ensure long-term sustainability of these projects. Skills deficits were evident, particularly in specialised technical areas. The digital field also changes rapidly, which creates difficulties in maintaining technical infrastructures over time.

A key problem highlighted by many of our interviewees was the difficulty in preserving born-digital content. Digital storage systems are not particularly robust and degrade over time or become unusable as the technology evolves. Digital formats change and become inaccessible; in some cases, archival formats were not available. In many instances, particularly with respect to e-mail, it was not clear what digital content should be saved for future generations and what could be safely discarded. In the absence of clear guidelines, ad hoc and sometimes erratic decisions were being made.

There were concerns that it might be assumed that digitisation would replace or supersede the conservation of the physical objects, which should not be the case. Digital objects are harder to preserve over time than physical objects. The value of digital objects comes instead from their ability to be found and shared across collections and manipulated and interrogated by users. This value is severely limited, however, if collectors do

Digital objects are harder to preserve over time than physical objects. The value of digital objects comes instead from their ability to be found and shared across collections and manipulated and interrogated by users.

not apply the same metadata standards or if the data are contained in very different formats. This survey indicated that ensuring interoperability is a critical challenge for the Digital Repository of Ireland and its stakeholder community. The content that is in most danger is born-digital, content that is being created right now but is under threat of being lost to future generations (the 'digital black hole'). It is highly likely that some born-digital content is already lost and may be unretrievable.

A number of tensions were also evident. In the context of insufficient staffing and IT support, digitisation programmes were driven by the immediate value returned for funding received, with the selection of collections often on the basis of high profile rather than a systematic approach to meet the wider community's longer-term needs. Whole-scale digitisation of collections was not being considered; if anything, digitisation projects were being scaled down or placed on hold. There is a further resource-driven tension between imaging projects and digitising projects. Imaging creates digital images of objects; it does not include the extra steps required by digitisation, such as the addition of metadata, transcriptions and contextual data and the secure storage of the objects created. The decision to image has in some cases led to the need to catalogue objects retrospectively to enable their reuse.

In developing content management systems, although there was a desire for more online content, there was also a desire to maintain control over content use and reuse in many cases. If only in-house access to databases is offered, it is much easier to control how data can be used. Inevitably, putting data online entails a certain loss of control. Many institutions wished to enable open access to their data, yet some of their content had ethical or copyright restrictions attached to it that made open access impossible. Finally, although most interviewees spoke of their willingness to share either their metadata or their data, this willingness was matched by the necessity of maintaining a link to the data, as retaining ownership was important in terms of creating their own brand and obtaining further funding for their institution. The nature of this link, whether it is attribution, citation and/or exposure to analytical software, will need to be explored in further discussions. While maintaining their own identity and brand, many institutions were also participating in large-scale international aggregated digital archives and catalogues, such as Europeana,⁴² recognising the value of an international profile. In addition, a significant trend internationally towards sharing publicly generated data is evident, and as new copyright and ethical frameworks are developed, barriers that militate against sharing may lessen.

⁴² See <http://www.europeana.eu/portal/aboutus.html> (accessed on 21 August 2012).



A number of trends were evident. Many of the interviewees were increasingly considering open-source solutions to their data management problems, in part because open-source is more affordable but also to avoid future archival problems attached to using proprietary systems. Many of the interviewees had legacy systems in place yet wished to adopt new technology, particularly in terms of enabling greater use of the internet. A key challenge faced by many was how to go about integrating multiple different systems (for example, different catalogue record systems).

Many of the interviewees were forward-looking in seeking new ways to develop and enhance the resources under their care. Many were interested in developing online educational tools, with a few engaged in or expressing interest in creating collaborative digitisation programmes, such as crowd-sourced annotation, metadata generation, and transcription or imaging of content. A number of institutions were using social media (particularly Twitter and Facebook) to generate audience awareness and interest in their collections.

The Irish digital landscape is rich, and the range of data held is diverse and impressive. It encompasses music, films, radio and television programmes, manuscripts, maps, art, architectural drawings, correspondence, interviews, archaeological surveys, newspapers, diaries, images (including 3D) and new media, including user-generated content. In the absence of a national strategy for protecting our digital cultural and social heritage, these objects are in danger of being lost to future generations. This is the DRI's contention, and our goal for 2013 and 2014 is to work with the community in the development of national guidelines for digital preservation and access, in order to inform future policy for our cultural and social digital heritage.

In the absence of a national strategy for protecting our digital cultural and social heritage, these objects are in danger of being lost to future generations.

Appendix 1: Methodology

Stakeholder interview sample selection

An intrinsic part of information-gathering for both requirements and policy is conducting stakeholder interviews. By considering the policies and practices already in place, the DRI can develop in a way that is supportive of existing institutions and projects. The DRI potentially includes data types and collections from various cultural, national and independent institutions that all have very special (although some overlapping) requirements and policy concerns for data ingestion, functionality, preservation and access. Other content providers include private or individual researchers. In addition, the DRI must satisfy those who wish to use the system and its collections.

Organisations interviewed to date (2012)

All-Island Research Observatory, National University of Ireland Maynooth

An Foras Feasa, National University of Ireland Maynooth

Clare County Library

Crawford Gallery, Cork

Digital Humanities Observatory, Royal Irish Academy

Discovery Programme

Dublin City Archives

Dublin City University Library

Economic and Social Affairs Institute

Health Research Board

Hunt Museum, Limerick

Irish Architectural Archive

Irish Film Institute

Irish Manuscripts Commission

Irish Museum of Modern Art

Irish Qualitative Data Archive, National University of Ireland Maynooth

Irish Qualitative Data Archive/National Institute for Regional and Spatial Analysis, DRI
Demonstrator Project, National University of Ireland Maynooth
Irish Traditional Music Archive
National Archives of Ireland
National Centre for Technology in Education
National Folklore Collection, University College Dublin
National Gallery of Ireland
National Irish Visual Arts Library, National College of Art and Design
National Library of Ireland
National Museum of Ireland
National University of Ireland, Galway, DRI Demonstrator Project
National University of Ireland, Galway, Library
National University of Ireland Maynooth, DRI Demonstrator Project
National University of Ireland Maynooth, Library
Oral History Network of Ireland
Raidió na Gaeltachta
Royal Irish Academy
RTÉ Digital
RTÉ Sound Archive
School of History and Archives, University College Dublin
School of Information and Library Studies, and Earth Institute, University College Dublin
Trinity College Dublin, Library
University College Dublin, Library
University of Limerick, Library and Department of History

Ethics approval/consent form for user/stakeholder interview recordings and transcriptions




Respondent information sheet (in-depth qualitative interviews)

Digital Repository of Ireland (DRI): key consultations

Thank you for agreeing to participate in this study. The DRI research consortium is tasked with building a robust, scalable, accessible and sustainable trusted digital repository (TDR) and access repository for the humanities and qualitative social sciences. As part of our work towards meeting this goal, we are documenting the experiences, concerns and preferences of key stakeholders in the research, library and archiving process. Your contribution in this regard will be extremely valuable.

The research is being carried out at the National University of Ireland Maynooth.

The investigators are:

-  Dr Aileen O'Carroll, Policy Manager, DRI/National University of Ireland Maynooth
-  Dr Sharon Webb, Requirements Manager, DRI/National University of Ireland Maynooth
-  Dr Sandra Collins, Director, DRI/Royal Irish Academy

With your permission, the interview will be recorded. Afterwards it will be written up/transcribed. Both the recording and the interview notes/transcription will be stored in a locked cabinet in the project head office at NUI Maynooth.

Once all the interviews are completed, the audio recordings and interview notes/transcripts will be deposited in an archive, where other bona fide researchers may consult them. You may be happy to be personally identified in these public materials. However, if you wish, your name will be removed, and your comments made unattributable.




Once again, we thank you for your participation. However, it is important for you to know that your participation in the research is entirely voluntary. You may withdraw your consent to participate at any time, without obligation.

Having read this information sheet, please read and sign the consent form.

Consent form (in-depth qualitative interviews)

Project Title: Digital Repository of Ireland (DRI): key consultations

The investigators are:

-  Dr Aileen O'Carroll
-  Dr Sharon Webb
-  Dr Sandra Collins

Material gathered during this research will be treated as confidential and securely stored in a locked cabinet at NUI Maynooth. You have the right to access any of your interview materials (tapes, transcripts and notes) at any time.

Please answer each statement below concerning the collection of the research data.

1.	I have read and understood the information sheet.	Yes 	<input type="checkbox"/>
		No 	<input type="checkbox"/>
2.	I have been given the opportunity to ask questions about the study.	Yes 	<input type="checkbox"/>
		No 	<input type="checkbox"/>
3.	I have had my questions answered satisfactorily.	Yes 	<input type="checkbox"/>
		No 	<input type="checkbox"/>
4.	I understand that I can withdraw from the study at any time without having to give an explanation.	Yes 	<input type="checkbox"/>
		No 	<input type="checkbox"/>
5.	I agree to the interview being audiotaped and to its contents being used for research purposes.	Yes 	<input type="checkbox"/>
		No 	<input type="checkbox"/>

Below are sets of statements that give you, the interviewee, a series of options about how you wish your interview to be used. Please answer each statement.

6. I agree to being identified in this interview and in any subsequent publications or use.

Yes

☐

No

☐

If you answered 'Yes' to Q. 6, please go directly to Q.8

If you answered 'No' to Q. 6, please also answer Q.7

7. Where used, my name must be removed and my comments made unattributable.

Yes

☐

No

☐

8. I agree to the interview notes/transcripts (in line with the conditions outlined above) being archived and used by other bona fide researchers.

Yes

☐

No

☐

9. I agree to my audiotapes (in line with the conditions outlined above) being archived and used by other bona fide researchers.

Yes

☐

No

☐

10. I would like my name acknowledged in the report and on the project website (without linking it to content or quotation).

Yes

☐

No

☐

Name (printed) _____

Signature _____ Date _____

Your contribution is greatly appreciated. Feel free to contact us if you have any further questions.

Dr Aileen O'Carroll: Phone: (01) 708 3596/E-mail: aileen.ocarroll@nuim.ie

Dr Sharon Webb: Phone: (01) 708 7182/E-mail: sharon.webb@nuim.ie

Dr Sandra Collins: Phone: (01) 609 0668/E-mail: s.collins@ria.ie






If during your participation in this study you feel the information and guidelines that you were given have been neglected or disregarded in any way, or if you are unhappy about the process, please contact the Secretary of the National University of Ireland Maynooth, Ethics Committee at pgdean@nuim.ie or (01) 708 6018. Please be assured that your concerns will be dealt with in a sensitive manner.

Topic guide for user/stakeholder interviews

The key approach is to use open-ended questions (e.g., can you tell me about/can you describe?), following the flow of the interviewee and only directing if the issues that need to be discussed do not emerge naturally in the course of the conversation.

We consider the resource/archive in terms of its current data lifecycle. Our aim is to establish how users/stakeholders currently support their digital resources/objects and how they develop and maintain their data archives/repositories. This will assist the DRI in setting key objectives and priorities.

STAGES:

- Pre-ingest:  The activities surrounding the data before they are prepared for archiving.
- Ingest:  Preparation and deposit of data into archive.
- Preservation:  Fulfilling the archive's responsibility to preserve data.
- Dissemination/Reuse:  Fulfilling the archive's responsibility to enable dissemination/reuse of data.
- DRI:  Future development in a federated repository.

KEY TOPICS

QUESTIONS

Stage in archive lifecycle: Pre-ingest

Digital objects/resources. Quantity, data formats (.txt, .doc), processes of digitisation (crowd-sourcing)

Computer or software systems in use

User interfaces (bespoke, particular product)

Static or living archive

Bilingual data

Can you tell me about your resource/archive/repository?

Can you describe your data/content?

Are all your data digitised?

Can you describe the digitising process?

Can you describe the current system you use for your data collection?

How do you envisage your resource developing in the future?

Expected data growth (in a two-year period, for example)

How much data are there now for ingest into the DRI?

Data quality assessment/quality control process (in terms of data formats and data content)

How do you assess data/content quality?

Stage in archive lifecycle: Ingest

Nature of data (special concerns, sensitive data, rarity, commercial issues). Access issues/policy

In terms of archiving or storing your data, are there any particular concerns or considerations? How did you address them?

KEY TOPICS

QUESTIONS

Ownership/copyright

Who owns the data? Are there copyright issues? Do you have licensing agreements?

Intellectual property

Are there any IP issues?

Collection priorities

How do you source the data?

Do you have special priorities?

Catalogue ontology/thesaurus

Have you developed a catalogue? If so, can you describe it?

Metadata formats

What metadata standards do you use?

Database formats

Do you know what database system you are using (MySQL, Excel, XML etc.)?

Linked data

Open data

Stage in archive lifecycle: Preservation

Future-proofing; data formats/longevity of data

Can you describe your preservation process, if any?

Data security (physical threats, virtual threats)/redundancy

Where are the data physically stored?

What security systems do you have in place, if any?

What level of redundancy (people, software, hardware, organisations) would you like to see implemented so that you feel the DRI is trusted, e.g., is two copies of data safe, three copies?

Do we need more than one person who is able to develop/maintain the system?

KEY TOPICS

QUESTIONS

Budgets

Do you include budget lines for preservation/ingest/storing data in a repository in your proposals?

Are you thinking about it?

Stage in archive lifecycle: Dissemination/Reuse

User experience/expectations (students, researchers etc.)

Can you describe who uses your data?

How do you see users in the future?

User tools

Do you provide any tools to enable the user to interact with the data?

Address concerns surrounding data security

How is the DRI 'trusted'?

DRI/organisation and infrastructure

Expectations of the DRI

What is important to you in terms of developing your resource?

Scale of investment to date in the project and predicted future investment

What are your biggest challenges?

Appendix 2: Resources

Formats⁴³

Name	Resource
Formats: general	http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf ; http://www.nationalarchives.gov.uk/documents/selecting-storage-media.pdf
AIFF (Audio Interchange File Format)	http://www.digitalpreservation.gov/formats/fdd/fdd000005.shtml
CAD (Computer Aided Design)	http://www.trixsystems.com/cad.html
DCP (Digital Cinema Package)	http://indieranch.com/Post/DCP_master.html
DGN (Digital Negative Format)	http://www.fileinfo.com/extension/dgn
DjVu	http://djvu.org/resources/
FLAC (Free Lossless Audio Codec)	http://flac.sourceforge.net/faq.html
JPEG (from 'Joint Photographic Experts Group')	http://www.jpeg.org/faq.phtml
JPEG 2000	http://www.jpeg.org/faq.phtml?action=show_answer&question_id=q3d5bc0701c9b6
Microsoft Office	http://office.microsoft.com/en-ie/support/?CTT=97
Microsoft Word	http://office.microsoft.com/en-us/word-help/
MP3	http://telos-systems.com/techtalk/hosted/Brandenburg_mp3_aac.pdf
MPEG (from 'Moving Pictures Experts Group')	http://telos-systems.com/techtalk/hosted/Brandenburg_mp3_aac.pdf
MPEG-2, MPEG-3, MPEG-4	http://telos-systems.com/techtalk/hosted/Brandenburg_mp3_aac.pdf
Open Office	http://incubator.apache.org/openofficeorg/index.html
PDF (Portable Document Format)	http://www.adobe.com/products/acrobat/adobepdf.html
PDF/A (Portable Document Format, Archive Standard)	http://www.adobe.com/enterprise/standards/pdfa/
QuickTime	http://www.apple.com/quicktime/what-is/
RAW	http://www.rondayphotography.com/Understanding%20the%20RAW%20File%20Format.htm
RTF (Rich Text Format)	http://office.microsoft.com/en-us/word-help/about-rich-text-format-documents-HP001004477.aspx

⁴³ We would like to thank DRI student interns Sam McGrath and Donal Fallon for their assistance in preparing this appendix.

Name	Resource
SVG (Scalable Vector Graphics)	http://www.w3.org/Graphics/SVG/
TEI (Text Encoding Initiative)	http://www.tei-c.org/index.xml
TIFF (Tagged Image File Format)	http://www.awaresystems.be/imaging/tiff/faq.html
WAV (Waveform Audio File)	http://www.digitalpreservation.gov/formats/fdd/fdd000001.shtml
WCS (Web Coverage Service)	http://www.opengeospatial.org/standards/wcs/
WFS (Web Feature Service)	http://www.ogcnetwork.net/wfstutorial
WMS (Web Mapping Service)	http://www.opengeospatial.org/standards/wms/
WordPerfect	http://www.corel.com/corel/index.jsp
WordStar	http://www.wordstar.org/index.php/wordstar-history
XML (Extensible Markup Language)	http://www.w3.org/XML/

Databases and spreadsheets

Basis	http://www.basis.com/database-management
CSV	http://www.imf.org/external/help/csv.htm
eXist-db	http://exist-db.org/exist/credits.xml
FileMaker Pro	http://www.filemaker.com/company/
Microsoft Access	http://office.microsoft.com/en-us/access/what-is-microsoft-access-database-software-and-applications-FX102473444.aspx
Microsoft Excel	http://spreadsheets.about.com/od/tipsandfaqs/f/excel_use.htm
MySQL	http://www.mysql.com/industry/faq/
Oracle	http://www.orafaq.com/wiki/Oracle_database_FAQ
PostgreSQL	http://www.postgresql.org/docs/faq/
RDMS	http://searchsqlserver.techtarget.com/definition/relational-database-management-system
SQL Server	http://www.microsoft.com/sqlserver/en/us/product-info.aspx

Digital asset/library/content management systems

Adlib	http://www.adlibsoft.com/support/faqs
ArtBase	http://www.artbaseinc.com/faq.swf
Drupal	http://www.3dissue.com/forums/forum/3d-issue-knowledge-base/

Name	Resource
DSpace	http://libraries.mit.edu/dspace-mit/about/faq.html#what
EPrints	http://www.eprints.org/openaccess/
eMuseumPlus	http://www.zetcom.com/products/emuseumplus/?no_cache=1&sword_list[0]=museum
ExpressionEngine	http://expressionengine.com/overview
Fedora Commons	http://www.fedora-commons.org/about
Fez	http://fez.library.uq.edu.au/wiki/Main_Page
FileMaker Pro	http://www.filemaker.com/company/
Kentico	http://www.kentico.com/Company
MySQL	http://www.mysql.com/industry/faq/
Oracle	http://www.orafaq.com/wiki/Oracle_database_FAQ
PostgreSQL	http://www.postgresql.org/docs/faq/
RDMS	http://searchsqlserver.techtarget.com/definition/relational-database-management-system
SQL Server	http://www.microsoft.com/sqlserver/en/us/product-info.aspx
TYPO3	http://typo3.org/about/
VTLS VITAL	http://vitalusers.wikidot.com/
WordPress	http://wordpress.org/about/

Metadata and vocabularies

AARC2	http://www.aacr2.org/about.html
Art & Architecture (Getty)	http://www.getty.edu/research/tools/vocabularies/aat/about.html
DDI	http://libraries.mit.edu/guides/subjects/data/archiving/ddi.html
Dublin Core	http://dublincore.org/specifications/
Dewey Decimal Classification	http://www.oclc.org/dewey/about/default.htm
EBU Core	http://tech.ebu.ch/lang/en/MetadataEbuCore
ESE	http://www.europeana.eu/schemas/ese/
INSPIRE	http://www.esri.com/software/arcgis/arcgis-for-inspire/common-questions
IEEE LOM	http://ltsc.ieee.org/wg12/ ; http://www.dcc.ac.uk/resources/curation-reference-manual/completed-chapters/learning-object-metadata
IPTC	http://www.iptc.org/cms/site/index.html?channel=CH0099
ISAD(G)	http://archiveshub.ac.uk/isadg/

Name	Resource
MARC 21	http://www.loc.gov/marc/faq.html#definition
METS	http://www.loc.gov/standards/mets/METSOverview.v2.html
MODS	http://www.loc.gov/standards/mods/mods-overview.html
NISO MIX	http://www.loc.gov/standards/mix/
RDF	http://www.w3.org/RDF/
SPECTRUM	http://www.pro.rcipchin.gc.ca/GetForumRecord.do?type=sd&lang=en&id=FORUM_23068&ens=cnRsTGFuZz1lbiZydGxUeXBIPXNk
VRA	http://www.vraweb.org/about/index.html

User tools

3D Issue	http://www.3dissue.com/forums/forum/3d-issue-knowledge-base/
eMuseumPlus	http://www.zetcom.com/products/emuseumplus/
Facebook	https://www.facebook.com/help/?page=160699464027593&ref=hcsubnav
Flickr	http://www.flickr.com/about/
Google Maps	http://www.makeuseof.com/tag/technology-explained-google-maps-work/
InstantAtlas	http://www.instantatlas.com/support.xhtml
OpenStreetMap	http://wiki.openstreetmap.org/wiki/Main_Page
OpenZoom	http://www.openzoom.org/ ('No longer maintained or supported')
Sibelius Scorch	http://www.sibelius.com/products/scorch/index.html
Solr	http://lucene.apache.org/solr/
Tableau	http://www.tableausoftware.com/about
TimeMap	http://www.timemap.net/index.php
Twitter	https://twitter.com/about
Zoomify	http://www.zoomify.com/about.htm

All accessed August 2012

Appendix 3: Stakeholder Advisory Group members

The DRI is honoured to have such an expert group to whom to present our progress. The members of this group are as follows:

Name	Role/institution
Mr Nicholas Carolan	Director, Irish Traditional Music Archive
Dr Mary Clark	City Archivist, Dublin City Archives
Ms Patricia Clarke	Senior Policy Analyst, Health Research Board
Ms Fionnuala Croke	Director, Chester Beatty Library
Ms Catriona Crowe	Head of Special Projects, National Archives of Ireland
Mr John Fitzgerald	Librarian and Director of Information Services, University College Cork Library
Mr Chris Flynn	Principal Officer, Cultural Policy, Department of Arts, Heritage and the Gaeltacht
Mr Jonathan Grimes	Information and Digital Services Manager, Contemporary Music Centre
Dr Cathy Hayes	Administrator, Irish Manuscripts Commission
Dr John Howard	University Librarian, University College Dublin Library
Mr Olivier Kazmierczak	ICT Manager, National Museum of Ireland
Ms Beatrice Kelly	Head of Policy & Research, The Heritage Council
Ms Christina Kennedy	Senior Curator, Head of Collection, Irish Museum of Modern Art
Ms Múirne Laffan	Managing Director, RTÉ Digital
Dr Kevin Marshall	Head of Education, Microsoft Ireland
Dr Jason McElligott	Keeper, Marsh Library
Ms Kasandra O'Connell	Head of the Irish Film Archive, Irish Film Institute
Dr Catherine O'Connor	Director and Founding Member, Oral History Network of Ireland
Ms Gobnait O'Riordan	Director, Glucksman Library, University of Limerick
Mr Colum O'Riordan	Archive Administrator, Irish Architectural Archive
Mr Seán Rainbird/ Ms Andrea Lydon	Director, National Gallery of Ireland/Head of Library and Archives, National Gallery of Ireland
Ms Fiona Ross	Director, National Library of Ireland
Mr Paul Sheehan	Director, Library Services, Dublin City University Library
Ms Marie Wallace	Social Analytics Strategist, IBM
Dr Manus Ward	Scientific Programme Manager, Science Foundation Ireland



Digital Archiving in Ireland

National Survey of the Humanities and Social Sciences

Aileen O'Carroll and Sharon Webb

First published in 2012 by the National University of Ireland
Maynooth, Maynooth, Co Kildare.

© National University of Ireland Maynooth

When citing this report, please use the following wording:

O'Carroll, A. and Webb, S. (2012), *Digital archiving in Ireland: national survey of the humanities and social sciences*. National University of Ireland Maynooth. DOI: 10.3318/DRI.2012.1

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

ISBN 978-0-956326-76-8

Design and layout by Fidelma Slattery

Printed in Ireland by Walsh Colour Print

Contents

5	Director's foreword
6	The Digital Repository of Ireland
7	Executive summary
10	Introduction
12	Methodology
15	Types of digital data in Ireland
18	Digital preservation: 'sustainable access'
27	Digital file formats
32	Metadata and vocabularies: describing data
37	User tools
44	Structuring content: database formats/systems, content management systems and repository software
47	Absences
48	Conclusion
51	Appendix 1: Methodology
60	Appendix 2: Resources
64	Appendix 3: Stakeholder Advisory Group members

Director's foreword

The Digital Repository of Ireland (DRI) is conducting a national programme of stakeholder interviews to determine the digital preservation and access practices in cultural institutions, libraries, higher-education institutions, funding agencies and more. Our findings shape our requirements specification in building the national repository, and they are also the beginning of a process to agree national guidelines on digital preservation for the humanities



DR SANDRA COLLINS

and social sciences. Our approach is first to determine national practice, then to work with the community in building national guidelines and hence to inform national policy.

This report presents our first findings. It is important to share our experiences, to learn from one another and from best practice both nationally and internationally, in order to serve our community of users now and into the future. Community engagement and informed dialogue are an essential part of this.

I would like to convey my sincere thanks both to the authors of this report and to the 40 respondent institutions that gave their time and support so generously towards this goal.

Much work remains to be done; further engagement is under way with our wide circle of stakeholders, and ascertaining digital practices is only the first step towards national guidelines that will be designed with the community, to be used by the community. It is a challenging task and not one that we undertake lightly, but it is essential and we must act now.

The DRI is working to raise awareness of the need for and benefits of digital preservation and open access, while respecting and acknowledging ownership, copyright, intellectual property rights, privacy and confidentiality. Digital preservation of our social and cultural heritage is imperative, and this is exactly what is at stake today, unless we act together.

Dr Sandra Collins

Director, Digital Repository of Ireland
Royal Irish Academy

Digital Repository of Ireland

The Digital Repository of Ireland (DRI) is building an interactive national trusted digital repository for contemporary and historical, social and cultural data held by Irish institutions. The DRI is linking together and preserving the rich data held by Irish institutions, providing a central internet access point and interactive multimedia tools for use by the public, students and scholars. The DRI is a national e-infrastructure for the future of education and research in the humanities and social sciences.

The DRI Research Consortium comprises the partners: the Royal Irish Academy (lead institute); the National University of Ireland Maynooth; Trinity College Dublin; the Dublin Institute of Technology; the National University of Ireland, Galway; and the National College of Art and Design. We are also collaborating with a network of academic, cultural, social and industry partners, including the National Library of Ireland, the National Archives of Ireland and Raidió Teilifís Éireann. We were awarded €5.2m from the Higher Education Authority's Programme for Research in Third-Level Institutions, Cycle 5 (funded as the 'National Audio Visual Repository'), and have also received awards from Science Foundation Ireland, Enterprise Ireland and the Ireland Funds.

...an interactive national trusted digital repository for contemporary and historical, social and cultural data held by Irish institutions.

Please visit our website, www.dri.ie, to learn more.



Executive summary



DR AILEEN O'CARROLL



DR SHARON WEBB

The Digital Repository of Ireland interviewed 40 institutions concerned with the humanities and social sciences about the procedures and practices that they have adopted in order to archive and care for the data in their collections. The interviews focused in particular on the care of digital data. The DRI will use the information generously provided to inform both the design and implementation of the national repository and the development of national guidelines, which will be designed with the community, for use by the community.

Our findings to date address different aspects of the digital lifecycle and are summarised below.

Types of digital data

A wide range of types of digital data were being cared for and created, including digitised manuscripts, photographs, moving images and audio material, as well as geospatial and geographical raw data. Digital data were generated by the digitisation of analogue material in collections, but increasingly data are created in digital form, that is, they are 'born-digital'. Most social scientific data, academic outputs and organisational data (including minutes of meetings, e-mails, webpages and social media) are now born-digital.

Sharing and reuse

There was an eagerness to enable sharing and reuse of digital data. Some collections, however, had copyright or ethical restrictions that limited these possibilities. There is a need for a national policy that would enable increased sharing and reuse of digital data.

Preservation

The preservation of digital objects was identified as a key challenge, a challenge that in many ways is more complex than the preservation of analogue objects. Digital preservation requires not only the secure storage of digital materials but also policies and workflows that ensure that such materials will be accessible and usable in the future. Although many interviewees were meeting the challenges of secure storage, few had workflows in place to ensure the successful migration of objects as current formats become redundant. Many of the interviews identified skill shortages and a lack of appropriate technical infrastructure as a key barrier to ensuring long-term preservation of digital objects. Born-digital data are in most danger of being lost to future generations.

Storage and formats

Although many institutions were able to store their current digital data, there were fears that, as the size of digital collections continued to grow, it would become difficult to afford and manage the storage space necessary. This was particularly the case in fields that cared for audio-visual files and 3D modelling files.

Data were stored in a relatively small number of formats. Institutions distinguished between preservation formats and access formats, often creating digital objects in both. Some formats that were appropriate to the data and in use internationally were not widely evident in Ireland. Policy guidance is required for those data types without archival formats in place and for format migration.

Metadata and interoperability

The added value of digital data over analogue data is that they enable institutions to share and build connections between collections. This is only possible, however, if standard metadata, fixed-word vocabularies and ontologies are used. Most of the metadata standards used were appropriate to the type of data that they were applied to; however, the DRI will face a critical and difficult challenge in ensuring that it is possible to build connections between datasets that are created using a range of metadata standards.

User tools

There is an increasing desire to ensure that today's users are able to interact with and add value to digital data. Many of the interviewees provided enhanced access either through curated collections and the provision of user tools, in particular geospatial mapping tools, which enabled users to manipulate data online, or through mobile applications that delivered content in new ways to users. A number used social media to raise the profile of objects in their collection and to generate new information about

the objects. These are welcome developments; however, they also raise the challenge of ensuring the sustainability of these user tools and the incorporation of user-generated content into existing collections.

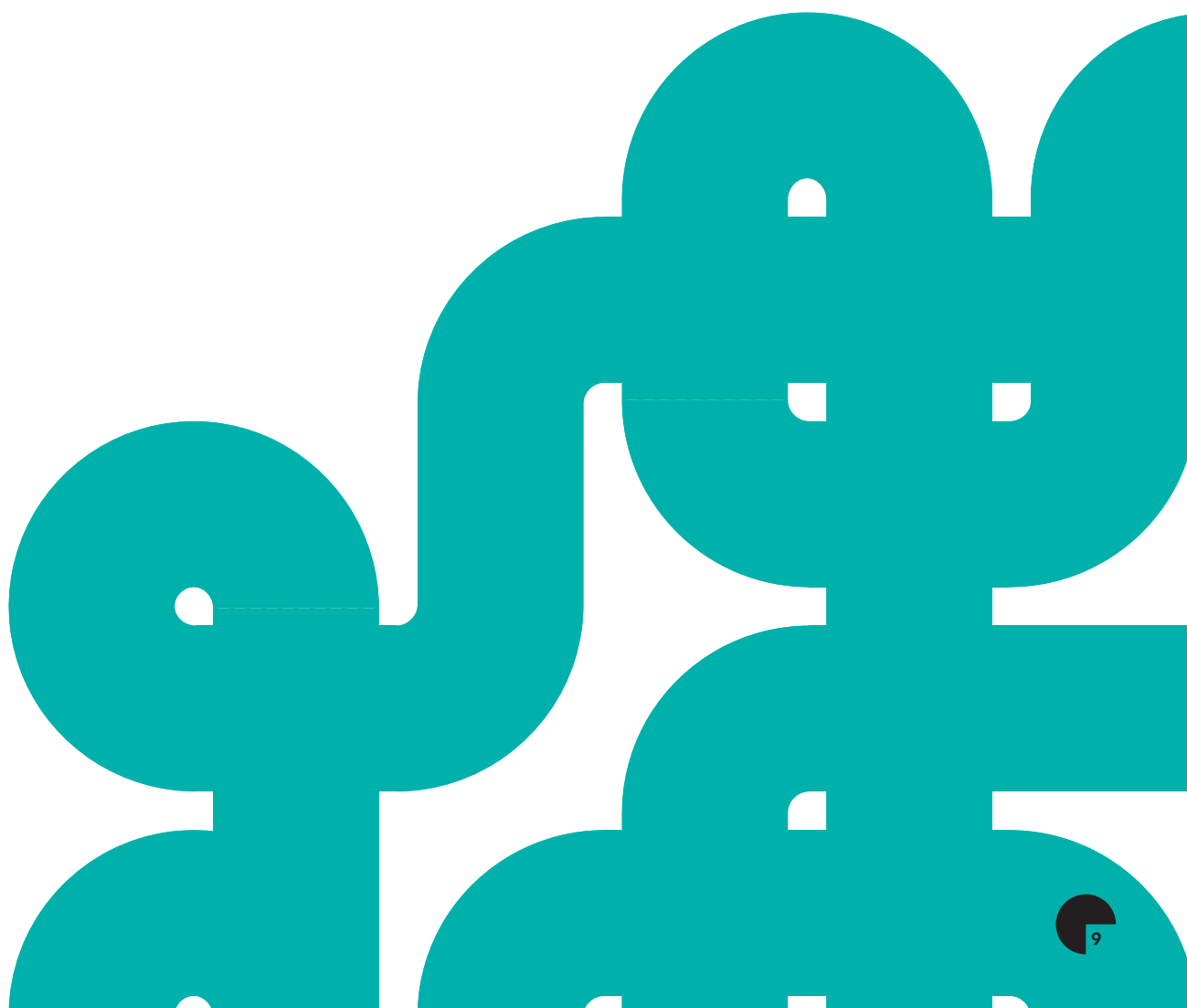
Structuring content

There are many database, repository and content management systems available. While MySQL was clearly the most popular database in use, there was surprisingly little consensus on the content management systems used, reflecting both the diverse needs of the community and the crowded nature of the content management field. Many institutions were upgrading or moving to new systems, in particular to ensure greater web access to their collections. This creates a challenge in integrating new and legacy systems.

Our findings reveal a vibrant and active community, which is cultivating and developing Ireland's digital landscape at a challenging time. The DRI will work with the community to develop digital guidelines and to provide preservation and access services to enhance current offerings.

Dr Aileen O'Carroll and Dr Sharon Webb

National University of Ireland Maynooth



Introduction

A trusted digital repository can be defined as a technical infrastructure ‘whose mission is to provide reliable, long-term access to managed digital resources for its designated community, now and in the future’.¹ The primary aim of the Digital Repository of Ireland (DRI) is to develop such an infrastructure. Yet, while we focus on the technical development of this system, a key issue is providing the associated services to the DRI’s ‘designated community’. This ‘designated community’ is diverse and is represented in the range of stakeholders with which the DRI has engaged to date. The successful implementation of the DRI’s goals and deliverables depends on the system’s ability to satisfy and implement the requirements of the DRI’s community. As part of this process, three members of the DRI research team, Dr Sandra Collins (Director), Dr Aileen O’Carroll (Policy Manager) and Dr Sharon Webb (Requirements Manager), carried out stakeholder interviews that surveyed the community’s current activities, requirements and desires in terms of digitisation, digital asset management, digital preservation and user engagement, as well as the challenges faced by, and the opportunities open to, this community.

These interviews are essential to the DRI’s ability to deliver a system that caters for the needs of its users. They inform the development of systems requirements and the DRI’s policy and usage guidelines. By incorporating requirements interviews into policy formulation and management, we aim to address the key concerns of the community and to develop strategies for digital rights management, digital preservation, access control and digital standards in response to those stated needs. The primary objective of these interviews is to ensure that the system is informed by authentic user requirements and that, as much as is possible, we support current good practices (e.g., data formats and metadata standards). This will allow us to build on the experience

The primary objective of these interviews is to ensure that the system is informed by authentic user requirements and that, as much as is possible, we support current good practices.

¹ *Trusted digital repositories: attributes and responsibilities*, an RLG–OCLC report (2002), available at <http://www.oclc.org/resources/research/activities/trustedrep/repositories.pdf> (accessed on 21 August 2012).

of the community in providing preservation and access services, while adding value and innovation to humanities and social science data through the DRI's infrastructure.

This report presents our findings from the initial phase of interviews and reveals that, although there is a diversity of interests and practices stemming from various perspectives, backgrounds and institutional obligations and remits, members of the DRI's designated community face many of the same issues and challenges in securing Ireland's digital cultural and social heritage. The report reveals a vibrant and active community that is cultivating and developing Ireland's digital landscape and provides essential information that will directly inform the DRI's development. This community is proactively unlocking Irish archives, providing access to content in new ways that add value to the material and transforming the user's experience.

This report also provides an opportunity for our stakeholders and interviewees to engage with this community. Along with the interviews, this report represents initial steps in terms of the DRI's stakeholder engagement. This dialogue will continue.

It is important that we acknowledge that this research builds on and complements previous Irish publications in this area, including the Library Council, *Our cultural heritage: building the gateway* (Dublin, 2004), the Irish Manuscripts Commission and the Digitisation Task Force, *Digitisation policy* (Dublin, 2007), the Spatial Heritage and Archaeology Research Environment IT, *A survey of digital practices in Irish archaeology* (Dublin, 2008), the Irish Qualitative Data Archive and the Tallaght West Childhood Development Initiative, *Best practice in archiving qualitative data* (Dublin, 2011), to name but a few. We would also like to acknowledge that this work builds on and complements recent and ongoing Irish initiatives. We envisage that this report will supplement that work and add to the enormous efforts already undertaken by our interviewees to develop and secure Ireland's digital future.

The DRI research team would like to thank everyone who has talked with us to date, and we look forward to engaging further with the community.

Methodology

The DRI conducted 40 requirements interviews with key stakeholders from December 2011 to August 2012. Figure 1 indicates the main spheres from which the interviewees were drawn and illustrates the range of stakeholders with which the DRI must interact, in terms of those who currently hold or produce digital content and those who use that content (there is, of course, overlap between the two categories; e.g., an academic researcher may both create research data and use data created by others). Individuals from the groups marked with an asterisk have been interviewed to date. The interview process will continue, to include those groups not interviewed in the first phase of consultation (see Appendix 1 for a list of organisations consulted).

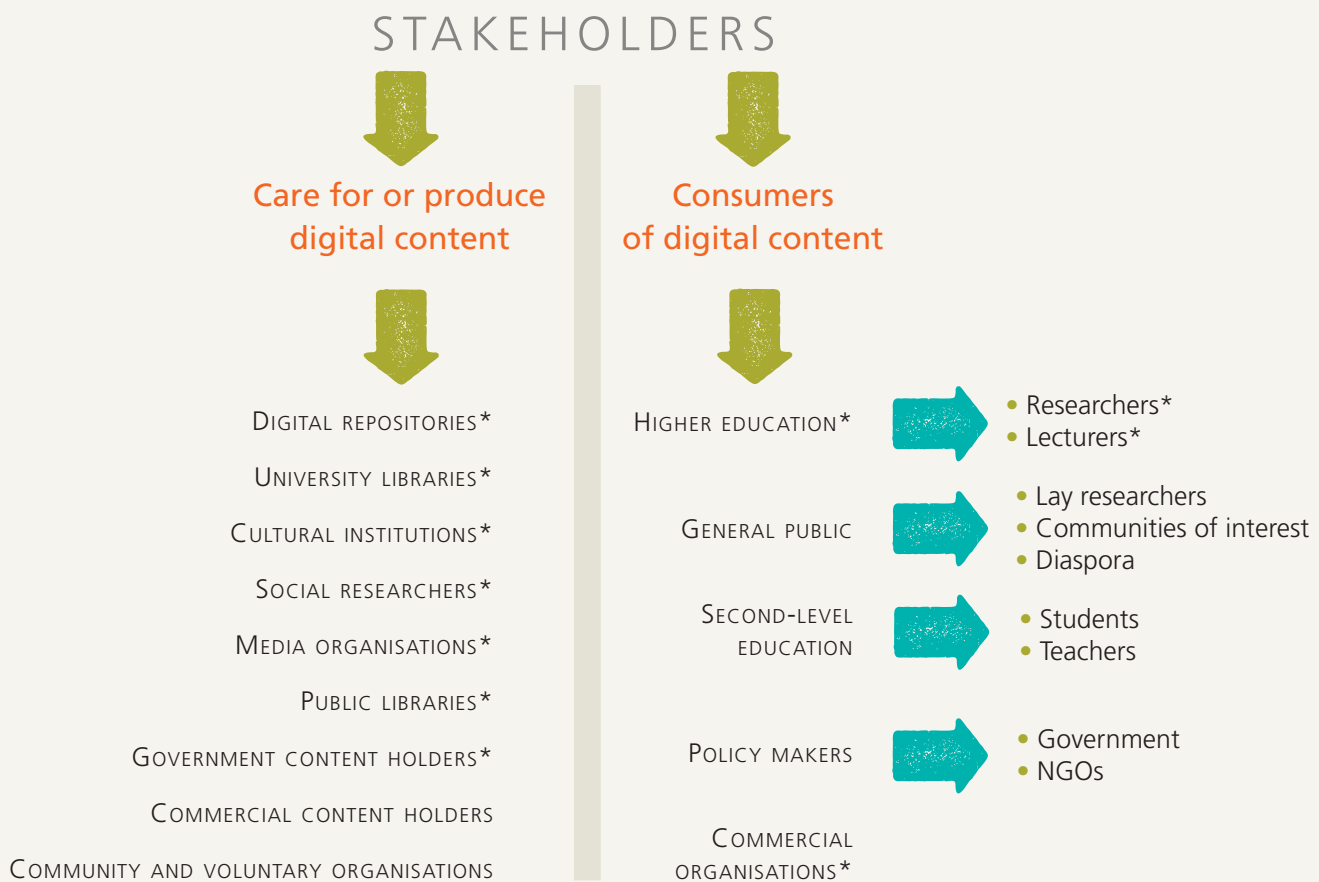


Fig. 1: Stakeholder interviews

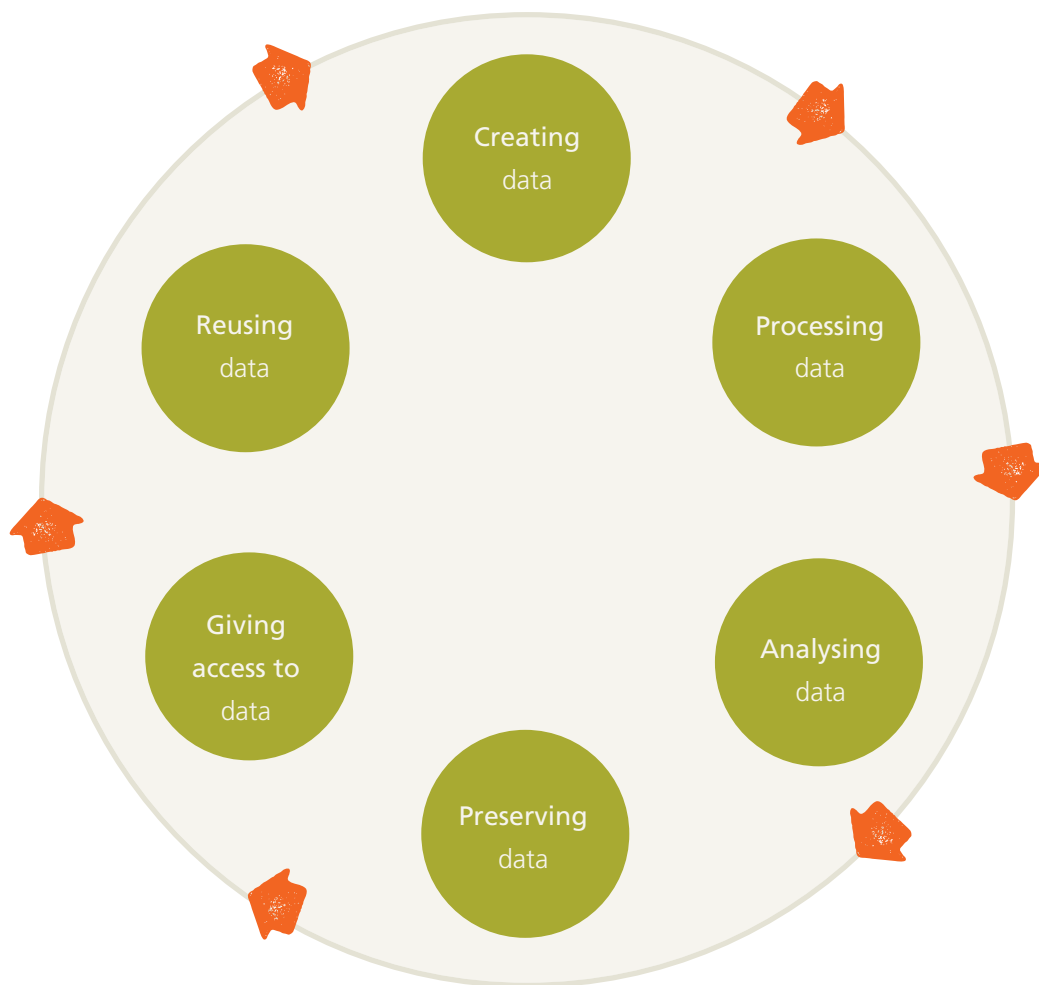


Fig. 2: Representation of the data lifecycle produced by the UK Data Archive²

'An inquiry, not an inquisition'³

Stakeholder and user interviews were structured but semi-formal, using a topic guide approach that allowed the interview to unfold as a free-flowing conversation while ensuring that all of the designated topics were discussed at some point. The topic guide ensured that the DRI interviewers received the desired information on a range of issues pertinent to requirements analysis and policy development, including data formats, metadata standards, existing systems, approaches to future challenges and expectations. We used the data lifecycle model developed by the UK Data Archive, detailing the phases of data creation, processing, analysing, preserving, access and reuse, as an initial topic guide template for the stakeholder interviews (Fig. 2). Pilot interviews were conducted with DRI partner organisations; the process was refined; and a final topic guide was developed that was used to inform further interviews (see Appendix 1).

² 'The Data lifecycle', UK Data Archive available at www.data-archive.ac.uk (accessed on 5 October 2012).

³ Wieggers, 2006, p. 57.

Ethics approval

The interviews were recorded (audio only) with the interviewees' permission. It is planned that these interviews will form a collection within the DRI and become part of the repository's project history. The interviews capture a moment in time after the initial phase of Ireland's digital archiving and before the wider, systemic approach being attempted through the DRI and therefore provide a view of Ireland's digital landscape at a critical juncture, highlighting the areas that are flourishing and those that are in need of support.

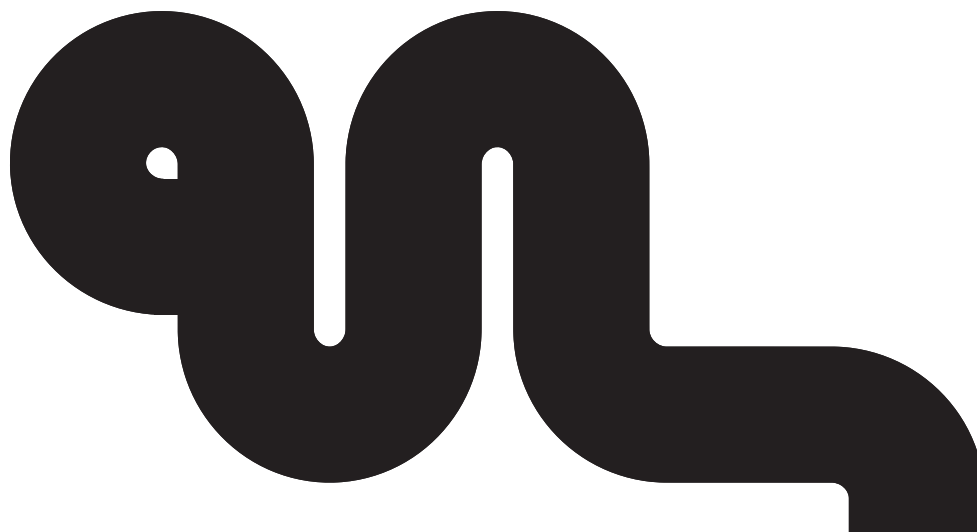
A copy of the consent form is provided in Appendix 1. Some interviewees requested that they not be identified, and consequently their responses are anonymised and unattributed in the text below.

Twenty-seven of the interviews have been transcribed and coded using the qualitative data analysis software MaxQDA. Encrypted WAV, RTF and Microsoft Word files are stored locally with the research team, and a back-up copy is stored unencrypted in a fireproof safe in a swipe-card-protected laboratory at the National University of Ireland Maynooth (NUI Maynooth). These files are accompanied by metadata using Dublin Core and Open Archives Initiative standards (see http://www.openarchives.org/OAI/2.0/oai_dc.xsd).

The interviews...provide a view of Ireland's digital landscape at a critical juncture, highlighting the areas that are flourishing and those that are in need of support.

Review process

This report was submitted for review to the DRI partners. After approval, it was submitted to the DRI Stakeholder Advisory Group (see Appendix 3). Feedback from both groups has been incorporated into this final report. Feedback included the suggestion that the DRI would conduct an audit of digitisation resources among the community to facilitate sharing of said resources. Additionally, a respondent indicated that a further survey to obtain more detailed information on the scale of data held by institutions and on technical infrastructures would be beneficial.



Types of digital data in Ireland

Of the 40 institutions interviewed, 36 were responsible for digital data. Because of the wide range of institutions, their relationship to the data varied markedly. Some were data creators, and some had a role in holding data for use by others. Others provided services that assist in the reuse and analysis of data that were also held elsewhere. The majority of the organisations interviewed held both analogue and digital data, digital data being a smaller, but growing, proportion of their collections. In terms of the types of digital data held, the range outlined below mirrors the variety of institutional roles.

-  Electronic text documents (transcripts, essays, diaries, theses, journal articles, books, journals, reports, learning resources)
-  Digitised images of analogue manuscripts, paintings, archival documents, newspaper clippings, printed ephemera
-  Photographs (including born-digital and digitised copies of prints and film negatives), some of which are historical and some contemporary, created in the research process or to document installations and other artworks
-  Moving images (including material produced for broadcast, home videos, born-digital and digitised copies of film and VHS, some including subtitles)
-  Interview and focus group audio files, home recordings
-  Radio programmes (born-digital and digitised copies of analogue radio, cassette tapes, records and cylinders)
-  Webpages, e-mails, social media, podcasts
-  Geospatial surveys, 3D documentation of objects, maps, architectural plans

Fig. 3: Types of digital data held by stakeholder groups

In the vast majority of cases, the collections were continuing to be added to, that is, they can be considered to be 'living archives' that develop and grow over time. Twenty-six of those interviewed had Irish-language data, and sixteen had data in other languages (such as Latin, French and Spanish).

'Born-digital' refers to material created in digital form; for example, an e-mail is a born-digital object. Thirteen of the institutions had born-digital material as part of their collections. For some, such as radio and television broadcasters and social research institutions, the transition from analogue to digital data is almost complete, with future data being generated and stored in digital form. Other institutions, such as art galleries, museums, and national and cultural archives, hold considerable analogue data. Here, digital data emerge from the institutions' own digitising processes (where progress depends on limited or reducing resources and funds). With all groups, some types of data, such as correspondence, are increasingly being generated in digital form, as e-mail replaces the letter. Other types of digital data, such as social media data (Flickr, YouTube, Twitter and Facebook), are being used in social scientific and humanities research by many of the institutions interviewed, but there is no concerted strategy to ensure that these data will be archived for the future.

Many of the interviewees were committed to allowing public access to their data. As Siobhán Fitzpatrick, Librarian at the Royal Irish Academy, explained:

[it] is a very important thing for us...the idea of having materials freely accessible as far as is humanly possible...Our overall access policy is to make material as freely available as possible...Most of the things that we would have done were paid for out of the public purse and our salaries are also paid by the taxpayer.

Copyright and confidentiality restrictions, however, limit the ability of institutions to share data. Copyright issues were of concern to many. Libraries were affected by the impact of copyright legalisation that placed access restrictions on the books, journals and collections that they held. Some institutions exercised copyright to generate revenue. Others exercised their copyright in order to limit unwanted reuse of their data; for example, one institution cited the reuse of a photograph in its collection by a commercial entity in a way that exposed the individuals in the photograph to ridicule. This type of misuse could be prevented by denying the right to reuse, although this requires that the institution be both aware of the reuse and in a position to defend its copyright.

Most social scientific data (and some donations to libraries and archives) had reuse restrictions placed on them that limited who would be able to access the data and

required that the anonymity of the interviewees be maintained. These limitations lessen over time: in 100 years, all data can be shared. Interviewees expressed concern, however, about the implications of such limitations for long-term preservation. The time and resources needed to ensure sustainable access to these objects, in order for them to become publicly available in the far future, had not been fully explored by any of our interviewees. It is, however, an area that requires immediate attention, and a number of institutions are developing pilot projects to address it.

In conclusion, institutions tasked with the care of data from the humanities and social sciences hold digital data in a wide variety of forms, ranging from the textual to the audio, the visual and the moving image. Although, in many cases, analogue collections are far greater than digital collections, digital data are increasing. In particular, social scientific research data, publishing and audio-visual data are increasingly created only in digital form. Digital data cannot be thought of as a simple copy of a physical object, and this report identifies many ways in which digital data differ from analogue data. Digital data create new possibilities but also new challenges. One of the possibilities is that there is a greater opportunity to share and reuse the data, such that the public has greater access to data held and produced by publicly funded institutions. This review found a marked interest in increasing access to digital data, including the use by many institutions of social media to engage with the public. However, there are important tensions. In the social sciences, where data are collected on the lives of contemporary individuals, a balance needed to be maintained between the rights of the public to access publicly funded data and the rights of research participants to have their confidentiality protected. Copyright brought an additional set of tensions that both restricted the sharing of data and protected the interests of individuals and institutions. While the copyright concerns attached to digital and physical objects are in very many ways similar, digital data carry additional opportunities and challenges. It is much easier to make collections and objects widely available by sharing them on the internet, but there was a clear sense that, once an object is released, it is extremely difficult, if not impossible, to police how that object might be used. Given that we are living in an increasingly digital world, there is a need for a national digital policy that capitalises on the possibilities of digital data and provides guidance on how to facilitate their sharing and reuse.

Given that we are living in an increasingly digital world, there is a need for a national digital policy that capitalises on the possibilities of digital data...

Digital preservation: 'sustainable access'

A major shift in the nature of living archives was evident, especially for those who dealt with modern content rather than solely historical material. Only one of our interviewees described its archive as static, that is, the institution did not actively receive or seek new analogue objects or material. Yet, with regard to digital collections, all of our interviewees had growing archives, indicating that the long-term preservation of digital data, as well as their capture, is a significant issue.

Digital preservation is therefore a challenge for all stakeholder institutions. Preservation is concerned with providing long-term access to digital objects, preserving continuity in form as well as functionality. It is not simply a back-up of data, because long-term digital preservation must consider format, software and hardware obsolescence, among other issues. Although it is possible for anyone to read a page from a book written 100 years ago, the same is not true of a floppy disk containing WordPerfect files from twenty years ago. Preservation is also resource-intensive and expensive, and all of our interviewees faced the challenge of providing long-term commitments to digital preservation, given current resource and funding restrictions and curtailments. The majority of archives, libraries, museums, universities, research institutes and other content owners from both the public and the private sphere that were interviewed had encountered difficulties with the preservation of digital data.

Although it is possible for anyone to read a page from a book written 100 years ago, the same is not true of a floppy disk containing WordPerfect files from twenty years ago.

After digitisation: 'custodianship of digital data'

Many institutions digitise on an ad hoc basis, either to meet user demands or for particular projects, for example, online exhibitions. One reason for this selection strategy is a lack of funding to digitise more comprehensively. Most telling, however, is that some view the creation of digital surrogates, at present, as an unreliable method of long-term preservation and see microfilm as a superior, tried-and-tested method. Yet, we must be clear: digitisation is not preservation. Institutions and funding agents need to consider the 'custodianship of digital data' after the digitisation process is complete, a step of

data management that is often overlooked in project funding. A number of our interviewees identified this problem, stating that although funding was allocated to digitise content, no allocation was made for the ‘custodianship of [the] digital data’⁴ generated through a funded project. While money was, and still is, allocated to generate digital surrogates of analogue objects, the long-term or even medium-term preservation of these digital objects was not accounted for as part of funding streams. This indicates that there is a significant funding, as well as methodological, gap between the generation of digital objects and the long-term preservation of content on completion of a project. This problem was also identified in research projects that produce various datasets during their lifetime, after which there are no contingencies for the long-term deposit of data or data management plans that would facilitate archiving and reuse. This problem could lead to duplication of research effort and costs.

Another view on the problem of digitisation is that while some funding may be available to create digital surrogates, there is little allocation to stabilise or conserve the original artefact. A number of institutions felt strongly that digital surrogates enhance access but should not be viewed as replacements. Yet digital preservation and access or dissemination of content are not mutually exclusive. Hugh Murphy, Senior Librarian, John Paul II Library, NUI Maynooth, commented that digital preservation facilitates ‘the user of today — the researcher of today’ but must also ensure the same level of quality and access for ‘researchers in twenty or thirty years’ time’.⁵

Born-digital data: ‘The [data] in most danger’

Born-digital data and archives pose more of a preservation problem than paper-based or physical objects and collections. Born-digital data are more complex than analogue data, as they encompass a multitude of data types, formats, applications and operating systems, as well as other hardware and software requirements and dependencies. One archivist stated:

in our case it is the born-digital [content] that is the big issue...that is, the [data] in most danger...We have no way of preserving it right now...the born-digital [data] that is being created right now or has been created over the last 30 years in so far as [what] survives...is the huge problem. And it is a problem everywhere; it is not just for us. But that is the one where we need to be looking for solutions.⁶

⁴ Anthony Corns, The Discovery Programme.

⁵ Hugh Murphy, John Paul II Library, NUI Maynooth.

⁶ Anonymous institution.

The major difficulty with born-digital content is that it is practically impossible to preserve or save it retrospectively. Digital data and media are fragile and volatile. There is therefore an immediacy of effort required to preserve long-term access to digital content. Una Walker, a post-doctoral researcher at the National College of Art and Design and a member of the DRI, shared her concerns on preservation of born-digital material: 'born-digital works...present particular problems in relation to preservation, partly because some [could] be networked, some [could] be using social media'.⁷ This indicates how born-digital works, in this instance new media artworks, are more complex than the digital surrogates of analogue material. Dr Walker is also closely involved with the National Irish Visual Arts Library, which has created a pilot project, the Digital Ephemera Archive, promoting the capture and preservation of digital ephemera.

The major difficulty with born-digital content is that it is practically impossible to preserve or save it retrospectively.

A small number of institutions informed us of pilot projects to capture web content based on a special remit such as a particular event or subject or to complement paper-based collections. The National Library of Ireland's born-digital collection includes 'web archiving activities around the 2011 General Election', hosted by the Internet Memory Foundation.⁸ The National Library of Ireland's website states that 'it is working towards collecting other born-digital material'.⁹ However, of the institutions and archives to whom we spoke, none indicated that it plans to web archive on a continuous or encompassing basis. Rather, web archiving is viewed as an activity related to particular projects and as such is restricted by limited resources and funding. As a result, very few web collections are being generated or maintained. However, organisations such as the Internet Archive, which offers services such as Archive-It, 'allow[ing] institutions to build and preserve their web archive of digital content',¹⁰ and the Internet Memory Foundation, as used by the National Library of Ireland, provide solutions to harness and gather institutional web content and could also be used to generate and host national web-based ephemera, material, data and collections. A limitation of these services is that they are hosted abroad, so that Irish institutions are depending on non-national, non-sovereign organisations to preserve the national digital heritage.

Another interviewee voiced concerns about what content is actually being captured, let alone preserved: 'there is a huge amount of digital material that is not being captured

⁷ Una Walker, National College of Art and Design.

⁸ National Library of Ireland, 'Born digital', available at <http://www.nli.ie/en/born-digital.aspx> (accessed on 1 August 2012).

⁹ *Ibid.*

¹⁰ See <http://archive.org/web/web.php> (accessed on 1 August 2012).

by anyone', a problem exacerbated by the fact that 'correspondence...and interaction between people' have changed.¹¹ Many institutions expressed particular concerns about the preservation of e-mail correspondence: it was not clear how such correspondence is being preserved and how the ethical issues associated with preserving and accessing it will be resolved. One respondent felt that people used e-mail, erroneously, as a record-keeping system to preserve their business records. Yet, without the policy-based use of e-mail archiving solutions or other software, at an institutional level there are no guarantees that important correspondence encapsulating institutional memory will be captured or preserved.

A number of archives, including the Irish Architectural Archive (IAA), already receive born-digital material and are having to face the problem of archiving and preserving digital collections that are unstructured, undocumented and more complex than their paper-based counterparts. The IAA is taking a very pragmatic and proactive approach to this problem and is looking to adapt existing technology to build a digital repository that can handle the complex issues associated with file formats (this is further discussed in 'Digital file formats', below). This development, it hopes, will reduce the loss of important organisational information and help the archive to grow its collections, pre-empting the surge of born-digital content while resolving the cataloguing and ingestion issues of said content. Colum O'Riordan, Archive Administrator at the IAA, hopes that the archive's problem of ingesting and accessioning born-digital material could be largely resolved, once practitioners and architectural practices use the repository to store current projects and adapt their workflows to allow access and repurposing of content.

Although changing people's workflows now will help with future accessions of born-digital collections, archivists currently face dealing with important digital collections that do not follow any consistent data management plan or approach. One archivist commented that the accession of a collection of CDs, representing an artist's life work, into the archive was 'classic of everything that is problematic with digital preservation'. The CDs contained:

[a] random file structure [with] random file names, all shortened to some sort of self-created code which evolves...from disk to disk [and a] very complicated directory structure with probably somewhere between 50 and 60 per cent of the directory...empty. Now, are they supposed to be empty, did he forget to fill them? We don't know.

¹¹ Irish Traditional Music Archive.

More problematic for digital preservation was the fact that no metadata or documentation accompanied the collection, and therefore basic, yet essential, information was missing, such as what programs were used and what versions. Fundamentally, the 'meta-data was zero'.¹²

Another archive received the office contents of an organisation that has ceased to operate, bringing ethical as well as technical difficulties: 'we took in a large collection of [the organisation's] archives and basically...it is in a complete jumble'.¹³ The archive received 'the contents of the hard drive of the computers' but has yet to catalogue the material and is faced with a new challenge, that of archiving a born-digital collection. This raises the question: in accepting a collection or archive, what level of responsibility for managing the data does the recipient assume? Kasandra O'Connell, Head of the Irish Film Archive (Irish Film Institute), spoke of a recent experience with a famous director who approached the IFI with floppy disks from which he could not retrieve the files, not only because of the media type but also because of the format. The IFI was able to recover the content, and the owner was able to see scripts that had not been seen in years.

The DRI research team asked interviewees about their preservation process and whether they had any written procedures or policies for future-proofing their content. A number of stakeholders indicated that they have in place or are in the process of developing digital preservation strategies or policy documents, and although the majority have yet to formalise their procedures, they are acutely aware of the implications and problems of digital preservation. In a few cases our interviewees had not previously considered written policies on preservation practices but said that the question raised an important issue that they were now prompted to address. One respondent felt that the DRI had a strong role to play in this area:

When I see the word[s] Digital Repository Ireland, I would expect to find born-digital records are stored there and preserved there so that they can be migrated forward into new formats and then preserved and made accessible at the right time. And I really think that is where the gap is more than any other gap.¹⁴

While archivists are often struggling to resolve these issues, there was also a clear sense that there is a lack of awareness of good digital preservation practice among the general

¹² Anonymous institution.

¹³ Anonymous institution.

¹⁴ Anonymous institution.

public. The IFI launched a short public appeal on YouTube about its preservation of Irish film records. Many of the suggestions documented on the YouTube site indicated a lack of understanding of digital preservation processes: 'all they have to do is go down to PC World and buy a few hard drives and then they can throw all the film out'; put it on YouTube and 'just get rid of copyright, that will preserve [it]'. In response, Kasandra O'Connell spoke of the Irish Film Archive's role in education and in increasing public awareness of digital preservation. She asserted that people need to be aware of the digital content that they personally hold and be proactive about its accessibility in the long term. She cited the example of photographs: 'Do you know where they are? Have you called them proper file names? Are you looking at them every couple of years?'. The IFI's experience indicates that there is a clear need for education and training on digital preservation. The lack of awareness among the general public is worrying, as it indicates, as do some of the experiences cited above, that much content is currently being lost.

Storage requirements

Directly linked to the challenges of digital preservation is the issue of digital storage. The storage requirements of our interviewees were diverse and ranged in size from 4 gigabytes for one archive to 65 terabytes for another. Despite the diversity in storage needs, which is linked to the type of content held by individual institutions, respondents faced similar storage issues and problems. Many informed us that they have to review their current storage procedures because of increasing demands. One described a common problem: 'the amount of space and storage needed for...digital material...is causing havoc with...IT departments'.¹⁵ These storage demands are linked to digitisation activities, an increase in the amount of born-digital data, including research data and academic outputs, and developments in

¹⁵ Anonymous institution.

media, as well as the adoption of increasingly complex data types. Additionally, as Brian Rice, Archivist at the RTÉ (Raidió Teilifís Éireann) Sound Archive, stated: 'storage... has become cheaper; standards and expected standards have increased'. These expectations impact on how data can be accessed, used and repurposed, as archives are faced with the problem of moving and processing data files that are growing substantially. The increased demands on storage reflect international trends and a global increase in the amount of data being produced and processed. The scale of this issue is highlighted by the fact that less than a decade ago 'there were only 5 exabytes of data online' but in 2010 'estimates put monthly Internet data flow at around 21 exabytes'.¹⁶ Our interviewees, therefore, are faced with the problem of ever-increasing storage demands, coupled with increased user expectations and the long-term preservation of massive datasets and files.

One institution estimated that one of its collections, which contained archival-standard TIFFs, was 40–50 terabytes.

As highlighted above, 'standards and expected standards have increased'. A number of interviewees spoke of imaging to archival standards to produce high-quality, high-resolution images and the consequent storage problems. As one respondent informed us:

We image...to the best of our ability to conservation standard, which is [a] 21 megapixel image. But [this] creates a massive storage issue...We have the conundrum as to whether or not we compress the files...We have to assess each project from the outset, and up until now we have done a conservation-standard image. But now we are finding...we are having to go back and resize everything, which is time-consuming, to compress it all.¹⁷

One institution estimated that one of its collections, which contained archival-standard TIFFs, was 40–50 terabytes. Given that the RTÉ Sound Archive estimated its current 'total requirement...[as] somewhere in the region of 65 terabytes', this storage requirement for a single collection is notable.

¹⁶ Audrey Watters, *The age of exabytes: tools and approaches for managing big data* (2010), available at http://www.readwriteweb.com/archives/download_our_latest_free_report_the_age_of_exabytes.php (accessed on 21 August 2012).

¹⁷ Anonymous institution.

Although this comparison between one project and an entire archive may be an extreme example, many interviewees were facing difficulties associated with massive datasets. Anthony Corns, GIS/IT Manager at the Discovery Programme, described how archaeological models and surveys generate huge datasets, citing one aerial image that consisted of raw data containing 150 million high points. He explained that when these raw data are subsequently modelled and rendered, the model could be at least 50 gigabytes. An additional challenge is how to render and make these data accessible over the internet.

Overall, our interviewees had a wide range of storage requirements and challenges. This diversity is directly linked to the various types of digital, and analogue, material that our stakeholders are concerned with; for instance, the IFI, which has yet to go 'down the route of digital preservation files' using DCPs (Digital Cinema Packages), informed us that a two-hour film of preservation quality would be approximately 6 terabytes. Another participant, dealing with high-resolution TIFFs, lower-resolution JPEGs and associated text encodings, estimated that it held just under 4 gigabytes of content for two digital projects but added that a third project, which includes a significant amount of moving images, would on its own require 5 gigabytes. The Irish Traditional Music Archive, which predominately deals with audio files, estimated that its servers held 7 terabytes of data but that it would require significantly more storage by the end of the year.

Most of our interviewees had growing digital collections and were looking for new storage solutions and ways to cope with increasing demands. A number informed us that they were at the limit of their storage capacity and were investigating alternatives to on-site storage. The RTE Sound Archive spoke about its investigations into corporate storage but concluded that 'the cost per terabyte and the pricing structure just made it impossible...to afford...[and other] options like cloud were just ridiculously expensive as well'.

Other interviewees also spoke of cloud storage, but only a handful actually used the cloud, for some, but not all, of their storage requirements. Events such as HEAnet's launch of EduStorage in April 2012 and the release of the Data Protection Commissioner's guidelines *Data protection 'in the cloud'* in July 2012 propose the use of virtual storage as a viable solution. Yet, although some institutions were considering the cloud for their storage solutions, many expressed reservations. One asked if it was trustworthy, and another librarian expressed concern about the potential loss of control of sensitive user information if cloud-based services were in different legal jurisdictions:

The conditions under which the American security forces could access your record would be different to what they'd get access to here if that

issue is to arise. So, for example, we have Chinese students studying here. Let's say the government back home wanted to find out who they were...we would just say no. We can't say what the Americans or the Israelis would say.¹⁸

Despite these reservations, one respondent stated that 'removable hardware, LTO tapes [etc.]...are no longer recommended for digital preservation'; instead, 'cloud computing' is advocated.¹⁹ In terms of back-up and disaster recovery, however, many institutions informed us that they still used these methods, among others.






In conclusion, this review indicated four crucial policy challenges. Firstly, digitising projects need to be undertaken in a framework that ensures long-term preservation of the digital assets produced; this includes explicit workflows across the data lifecycle. Secondly, of great importance for the DRI are the concerns raised by institutions about the difficulties in preserving born-digital content, concerns that highlight a policy gap and the need for guidelines, in addition to technical solutions. Thirdly, clear guidance, in terms of shared guidelines, appropriate workflows and preservation processes, is necessary to ensure that today's digital objects will be accessible to future generations. Finally, although many institutions are currently able to meet their storage requirements, these requirements have increased exponentially and will continue to grow.

¹⁸ Anonymous institution.

¹⁹ Anonymous institution.

Digital file formats

Digital data can be stored in a variety of formats; for example, there are over 60 common file formats that can be used to store electronic text.²⁰ These formats vary in many ways, but from an archivist's perspective the key issue is whether the format will be accessible in the future. As high-quality formats are larger in size, another issue is how to maintain a balance between archival demands of high quality and storage cost limitations. The US National Archive gives the following suggestions to those considering which format to use:

-  The format is publicly and openly documented.
-  The format is non-proprietary.
-  The format is in widespread use.
-  The format is self-documenting.
-  The format can be opened, read, and accessed using readily-available tools.²¹

In recognition of how quickly formats change, the US National Archive does not mandate the use of particular formats. Other archives, such as the UK Data Archive, take a similar approach. Rather than delineating the only formats that it is willing to handle, it outlines which formats it prefers and which are acceptable.²² This pragmatic approach is in recognition of the fact that, at times, archives have little choice in what formats they are offered: if a highly important sound recording is deposited in a low-quality format, the archive has little option but to accept it. Some institutions, however, generate content themselves, and here decisions need to be made about which form content should be generated in. As we have seen in 'Digital preservation', above, many of the organisations interviewed

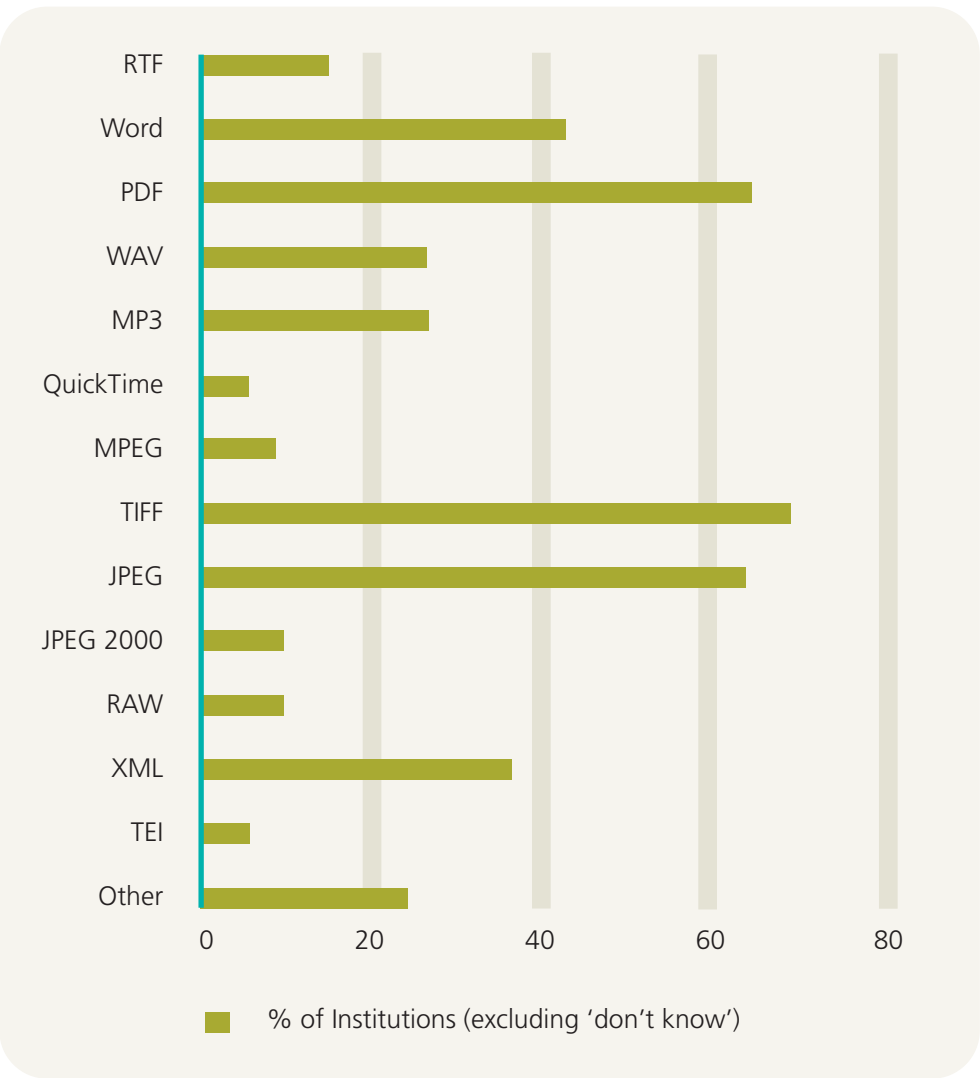
²⁰ Sources of further information on the formats mentioned in this section can be found in Appendix 2.

²¹ See <http://www.archives.gov/records-mgmt/initiatives/dav-faq.html> (accessed on 1 August 2012).

²² See <http://www.data-archive.ac.uk/create-manage/format/formats-table/> (accessed on 1 August 2012).

stored the same digital object in a variety of formats, a higher-quality format being used for preservation and a lower-quality format for sharing and access. Here, we outline which file formats the interviewees reported were held by their institutions.

Fig. 4: Formats used by institutions



From Fig. 4 it can be seen that the main format for textual data is PDF, with Microsoft Word in second place. Microsoft Word is a proprietary format, that is, it was created and is owned by a particular company (Microsoft) for use in its applications. The difficulty with proprietary formats is that should the company cease to trade, cease to support legacy software versions or change the nature of its software, it may be difficult or impossible in the future to access documents saved in that format: for example, documents written using the first word processor, WordStar, cannot be accessed on modern

versions of Microsoft Windows. As we saw earlier, one organisation still received documents in WordPerfect, a proprietary format that, since 2001, is available only for computers running Microsoft Windows (and OpenVMS). Many archives, such as the UK Data Archive, accept Microsoft Word documents, as they are so universally used, but are aware that this format may become obsolete over time.

Similarly, as PDF is so widely used, it is often accepted by archives; however, it is not considered an archival format. Although PDF was initially a proprietary format created by Adobe, it was released as an open standard in 2008. There are difficulties in using it as an archival format, however, because it does not embed the fonts that are used in the document. An archival version of the PDF format, PDF/A, was created in 2005 precisely to assist the long-term digital preservation of electronic documents. Microsoft Office and Open Office have introduced the ability to create PDF/A documents; however, they are not, as yet, widely used, and none of the institutions reported having PDF/A files. As with other archival formats, the files stored in PDF/A are much larger than those stored in PDF. In archival terms, RTF, while also a proprietary format, is preferable to Word and PDF, as it was created by Microsoft precisely to allow cross-platform sharing of documents and the specifications of the format have been published, allowing developers to include it in their software; therefore, it is far less likely to become unsupported in the future.

A number of the institutions that we interviewed held audio material, originating as broadcast material (RTÉ Sound Archive, RTÉ Raidió na Gaeltachta), field recordings of musicians (Irish Traditional Music Archive) or audio recordings of research interviews (Irish Qualitative Data Archive). Three audio formats were found in these collections: WAV, AIFF and MP3. WAV and AIFF are considered to be archival formats; however, the files that they produce are considerably larger than MP3 files (MP3 is not an archival format). WAV and BWAV are recommended for use in archiving by the International Association of Sound and Audiovisual Archives, while the UK Data Archive's preferred audio format is FLAC, but it considers either WAV or AIFF to be acceptable. If audio files have been donated in MP3 format, the archive may have no choice but to accept them. WAV files contain more information than MP3s, and it is therefore not meaningful to attempt to translate an MP3 file to WAV format. Indeed, one archive noted that when this was done (in error), an unwanted hum was added to the final files. AIFF, an audio format developed by Apple, was accepted by one of our interviewed organisations.

A number of the institutions held multimedia audio and visual formats. QuickTime files were held by NUI Galway Library and the IFI. QuickTime is a proprietary format created by Apple. MPEG-2, MPEG-3 and MPEG-4 formats were also used. MPEG is a standard

created by a working group of the International Standards Organisation and the International Electrotechnical Commission, and hence in archiving terms it is preferable to QuickTime. As with the audio formats above, however, institutions were not necessarily in a position to dictate which formats they would accept.

TIFF, JPEG, JPEG 2000 and RAW are all image formats. TIFF was the most popular format, used by 71% of the organisations interviewed. Files were imaged at 600 dpi, 400 dpi or 300 dpi. TIFF is considered to be the standard archiving format; however, TIFF files are considerably larger than JPEG files. As we have seen in 'Digital preservation' above, many archives produced JPEG versions for dissemination (particularly for use on the web) rather than preservation. JPEG 2000 was introduced in 2000 as an archival version of JPEG. It is not supported by many web browsers, however, and has failed to be widely adopted to date, although one of the institutions interviewed stored JPEG 2000 files.

A variety of organisations, including the Royal Irish Academy, the National Folklore Collection and NUI Maynooth Library, held RAW files. RAW files are the original image files created by digital cameras. They are a proprietary format, with each camera manufacturer creating its own version of RAW files. DGN (Digital Negative Format) is an open format created by Adobe with the aim of becoming a standard format into which RAW files could be converted for archival use. It is based on the TIFF format, although it has not yet been adopted as a standard. None of the institutions reported using DGN, and they may find it difficult in the future to access the RAW files that they hold.

Thirty-eight per cent of the organisations reported holding XML (Extensible Markup Language). XML was developed by W3C (World Wide Web Consortium) and is a non-proprietary, platform- and software-independent mark-up language used for data and document encoding, storage and transport. TEI (Text Encoding Initiative) is a set of XML-based guidelines for the digital encoding of literary and linguistic texts.

A number of organisations store formats that were special to their communities. The IFI holds DCP (Digital Cinema Package) data. DCP is a format created by a coalition of the major film studios to collate audio, image and data streams within a single format. The IAA holds CAD and BIM (Building Information Modelling) files, which document a range of digital information about the building process, from plans to manufacturers' details of a building component. The Discovery Programme, which creates 3D imaging of archaeological sites, stores 3D PDFs and raster data. It creates images using WMS (Web Map Service), a widely used format for transmitting geo-referenced map images over the internet that was created in 1999 by the Open Geospatial Consortium. Additionally, it uses WFS (Web Feature Service) and WCS (Web Coverage Service), which

are specifically designed to assist in the delivery of digital geospatial content over the web. Clare County Library has considerable map-based resources on its website.²³ These are stored as SVG files, an XML-based format for 2D vector graphics. Additionally, it stores maps in DjVu files, which have been created with DjVu image compression software. DjVu is an open file format and is used by the Internet Archive for its Million Book project. For those organisations that manage CAD or 3D modelling files, such as the IAA and the Discovery Programme, there are as yet no archival formats for these files. This problem is compounded by the use of proprietary software and further complicated by different software and application versions and releases. Colum O’Riordan, Archive Administrator at the IAA, spoke of the difficulty of dealing with CAD files, not only because there is no archival CAD format but also because backward compatibility is poorly supported. More challenging, perhaps, is the fact that the same CAD file opened in Microsoft Windows, Linux or Apple OS will not necessarily look the same. This is a serious problem for long-term preservation, especially because architects are now designing in 3D, moving away from CAD and into BIM, a move that will see architectural archivists and repositories dealing with the legacy of different design systems.

In conclusion, although many file formats exist, a relatively small number were in use in Ireland. There were commonalities in terms of the formats used for textual, image and audio data. The formats used for dealing with geospatial and broadcast material were not widely found across the institutions reviewed. This was a reflection of the uniqueness of this content in the Irish context; the formats used for these data were in common use internationally. Some archives (such as those dealing with 3D data or CAD files) faced a particular problem as preservation formats were not available. Although cost pressures could drive institutions to select non-preservation formats, in the main this was not evident. Instead, two types of format were often held: high-quality, large-size preservation formats and lower-quality, smaller-size formats for web dissemination. Those institutions whose remit was not primarily in the field of archiving (such as art galleries) sought guidelines on which formats would be most appropriate to preserve the data in their collections. A national preservation policy would also have to provide guidance on organisational policy and workflows to track the development of new formats and migration of data to new formats where necessary.

²³ See <http://www.clarelibrary.ie> (accessed on 2 August 2012).



Metadata and vocabularies: describing data

Metadata

Metadata are often described as 'data about data'.²⁴ The term refers to information attached to an object that tells us more about the object. A library catalogue record about a book is a form of metadata. Metadata can be documented in many different ways; metadata schema or standards are common or shared ways of documenting metadata. These schemas have emerged in different domains, as the key characteristics of the objects described vary according to the communities using the objects.²⁵ Shawn Day, Project Manager at the Digital Humanities Observatory, outlined the problem faced by many:

one of the biggest barriers we obviously face is just standards and formats. You are really dealing with the multitude of the ones out there. Part of this would have to do with...the fact that we are dealing with projects that have...been creating stuff over the past 15 years. Both within a world where these things are changing anyway but also in a situation where there has been no central authority by any stretch that people could consult to determine what are best practices.

In the digital world, metadata are important because if online collections use the same metadata schemas, it becomes possible to search across collections.

One-fifth of those interviewed were not aware of which schema they were using. The results in Fig. 5 are from those who reported their metadata use, and most institutions were using between one and four standards. Additionally, 6% of the institutions reported that they did not use any international standard but instead developed an in-house

In the digital world, metadata are important because if online collections use the same metadata schemas, it becomes possible to search across collections.

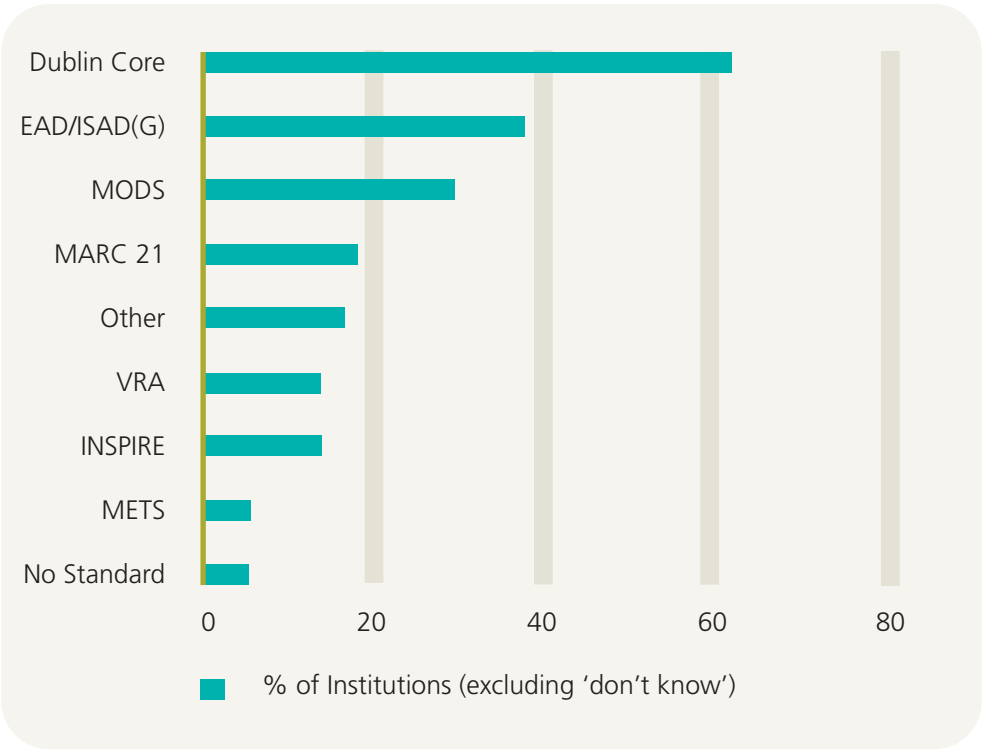
²⁴ Sources of further information on the metadata standards and vocabularies mentioned in this section can be found in Appendix 2.

²⁵ A metadata record, therefore, typically consists of a number of predefined elements representing specific attributes of an object, and each element can have one or more values; for example, most metadata schemas will have an element called <Title>, which refers to the name of the object that is being described.

description framework or were currently considering which standard to use. It is important to note that, once modified, 'standards' are no longer 'standards', which limits interoperability and thus the potential to share metadata between systems and organisations.

Dublin Core is one of the simplest metadata schemas ('Dublin' refers to Dublin, Ohio, which was the location of a workshop in 1995 from which the schema was developed). It consists of fifteen metadata elements, including title, creator, subject and description. As it is relatively versatile, it is widely used internationally. Indeed, it was the most popular metadata schema used by the interviewees, with 61% reporting that they used it. The next most popular metadata schema was ISAD(G), which was used by 39%, predominantly libraries, reflecting ISAD(G)'s (EAD) origins. This schema was developed for use in archives by the International Council of Archives, which published the first version in 1993. This schema contains 26 elements, of which six are mandatory: reference code, title, name of creator, dates of creation, extent of the unit of description and level of description.

Fig. 5: Metadata standards



The MARC format, used by 19% of the institutions interviewed, was created by the US Library of Congress in the early 1970s for use by libraries. The MODS standard is another initiative of the US Library of Congress and was established for use by libraries in describing bibliographic items. It was created as an alternative to the complexity of the MARC

format and the simplicity of Dublin Core. The MODS schema was used by 29% of the institutions. The METS schema, which was used in 6% of cases, also emerged from the US Library of Congress.

Metadata are field- or domain-driven, and therefore institutions that hold or use geospatial information are increasingly adopting INSPIRE, a standard for spatial information that was developed by the European Union in 2008. It was used by 13% of institutions, including the All-Island Research Observatory (AIRO), which provides map-based interfaces for government datasets, and the Discovery Programme, which curates 3D maps of archaeological sites.

Another standard that featured in some of the interviews was VRA, which was developed by the Visual Resources Association in 1996 to aid the description of visual objects. It was used by 13% of our interviewees, especially those, such as the National Gallery of Ireland, whose collections were made up largely of images.

The other standards in use (by a single institution in each case) were ESE, NISO MIX, IPTC, IEEE LOM, EBU Core and SPECTRUM. Two institutions mentioned adopting MODS in the future, and one mentioned future use of DDI (a standard applied to social science data). As we saw earlier, many collections contained digital data in a wide variety of languages. In most instances the metadata reflected the language of the original object (so, for example, an Irish-language song would have metadata in Irish).

Controlled vocabularies

Controlled vocabularies and thesauri are used in archiving to ensure that objects are described in common ways: e.g., a ‘worker’ could also be called an ‘employee’; the ‘labour market’ could also be described as the ‘job market’. For this reason, archivists have developed controlled vocabularies, thesauri and ontologies to give guidance to those adding data about an object (an ontology is a specification both of terms and of the relations between them).²⁶ Finding multiple instances of the same object becomes very difficult if that object is referred to in different ways on different records. For Irish data, a particular problem was noted in that multiple spellings were commonly used for many Irish surnames and place names (especially townlands). Damien Gallagher, at the time of the interview a software developer with An Foras Feasa, NUI Maynooth, faced this problem when importing an international database:

²⁶ For example, in the Getty Art & Architecture Thesaurus, ‘churches’ are a subcategory of ‘religious structures’, which are a subcategory of ‘single built works by function’.

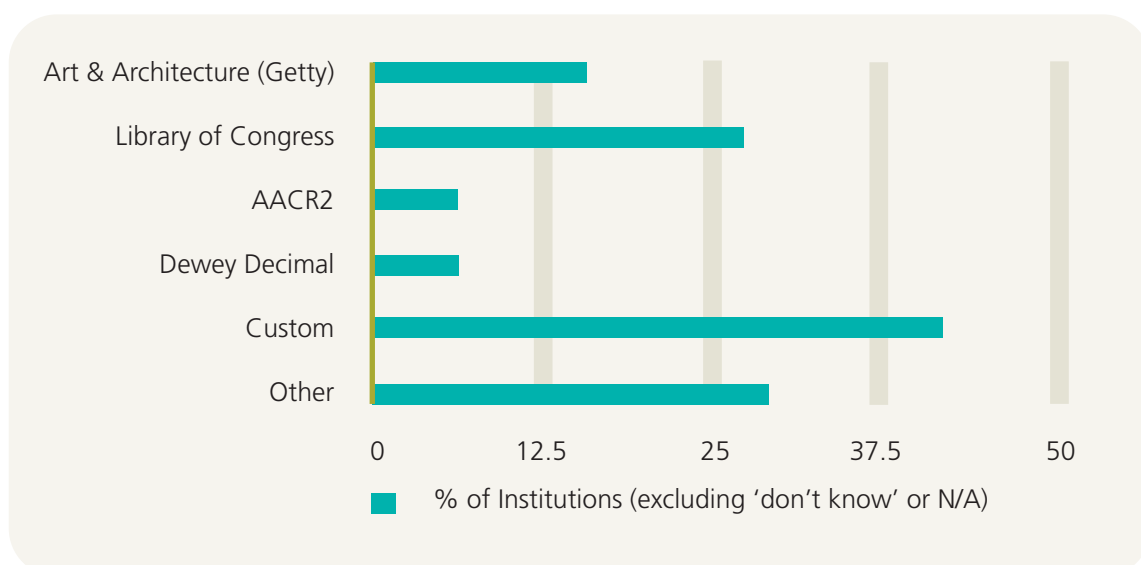
A lot of the records weren't written by the Irish people; they would have been written by French people or Spanish people, and they would have completely...misspelled [Irish names]. They would 'Frenchify' a certain thing. For example, I have seen Sweeney spelt in the French record as Suiney.

Críostóir Mac Cárthaigh, Archivist at the National Folklore Collection at University College Dublin, referring to a pilot project to index electronically the names, surnames and addresses of contributors to the National Folklore Collection, outlined the problem:

If you were searching for O'Sullivan, it would give you Ó Súilleabháin and the different various spellings of Súilleabháin. It was a big challenge as well because, in Ireland, no two people spell the townland the same.

Figure 6 below shows the considerable variation in the guides used, with 33% of institutions creating their own. The 'Other' category includes the following controlled vocabularies and guidelines: the UK Data Archive's HASSET Thesaurus, the Placenames Database of Ireland (www.logainm.ie), the Irish Public Service Thesaurus, the International Federation of Film Archives Taxonomies, the UK Archival Thesaurus, the Dictionary of Irish Biography and the British Architectural Library's Architectural Keywords. It should be noted that although some of these vocabularies are freely available, others are proprietary and require licensing for use, which can be expensive.

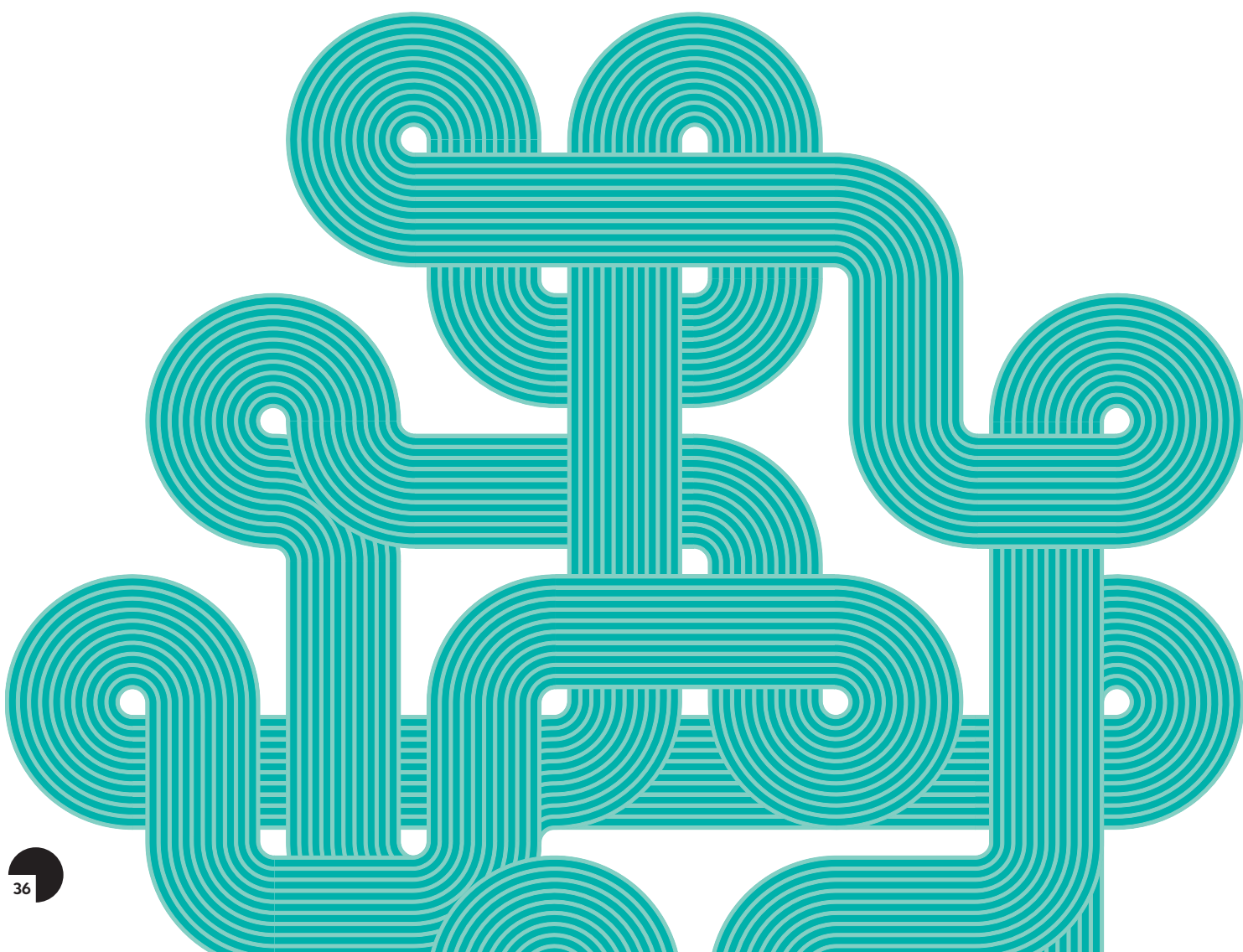
Fig. 6: Ontologies, thesauri and fixed-word vocabularies used



This review found that the majority of the institutions were using at least one of seven standards and that Dublin Core was by far the most popular. These standards were appropriate to the content and were also used by special user communities internationally; for example, libraries and archives

both in Ireland and internationally tend to use ISAD(G). A number of the institutions, however, used standards that were domain-special, e.g., INSPIRE, EBU Core and LOM, reflecting the uniqueness of the content that they held. This highlights the challenge faced by content holders in identifying which standards are most appropriate to their field: the temptation to develop one's own bespoke approach should be avoided, as it may limit future data connectivity. The challenge facing the DRI will be to facilitate interoperability between datasets developed in different domains. It is highly unlikely that the DRI will be able to support all of the metadata standards in current usage; however, it plans to support a suite of standards that balance best practice and common usage with current national investments.

The temptation to develop one's own bespoke approach should be avoided, as it may limit future data connectivity.



User tools

Collecting, managing and preserving digital material are crucial activities in the lifecycle of born-digital and digitised material. In order to engage users and to add value to these collections, however, many of our interviewees created and provided user tools to give audiences new ways to interact with digital material. These tools include user-generated annotations, crowd-sourced transcription correcting, networked mapping and graphing of relationships between material, text comparisons, data and geospatial visualisations, interactive maps, online exhibitions, interactive guides, interactive tables and educational tools. Mobile applications ('apps') were also developed, ranging from augmented reality to online catalogues and other visualisation aids for resource discovery, including timelines to narrow search criteria and results. Our interviewees were cognisant of the need to provide user tools, not only to provide new ways to access digital material but also to allow content to be reused, repurposed and reinterpreted, adding value to the content, as well as the archive. The development of user tools is in response to a noticeable shift in user expectations. Of this, one respondent stated that just using:

The development of user tools is in response to a noticeable shift in user expectations.

A PDF format...is not good enough anymore in terms of how people interact with [material]...[New] undergraduate[s] that come in...access and use the material in a very different form than perhaps what we created even three years ago. And how they expect it to be delivered, how they want to manipulate it, how they want to use it [have changed].²⁷

Providing new opportunities for users to engage with content also poses new challenges to content holders and tool developers, as they must grapple with the problem of providing sustainable user tools. This maintenance of functionality is another major issue for long-term digital preservation. Content holders must not only preserve digital objects but also provide long-term access to the context in which the objects are viewed,

²⁷ Anonymous institution.

visualised, used and manipulated. Copyright restrictions can also constrain what end-users can do with digital material, while funding and resource allocation to develop new user tools and resources is a major issue faced by all.

Exploring digital content: user tools for content engagement

The majority of our interviewees were identifying new ways to increase user engagement with the data in their collections. While access to material, both digital and analogue, is primarily achieved through the development of finding aids, 60% of our interviewees provided additional user tools to support resource discovery and enhance the user's experience with digital content and material. Although the remaining 40% did not yet provide any additional user tools, the majority were considering future developments in this area and a number discussed working with external partners.²⁸ Arlene Healy, Manager of Digital Systems and Services at Trinity College Dublin Library, identified user tools as an area for future development, stating that 'right now, our digital library displays the images after being browsed or searched so there is no...value added' to the objects, but she continued that 'it is definitely a future requirement'.

The integration of user tools into our interviewees' websites and systems include: in-house or bespoke development of applications; the use and customisation of open-source tools; tools developed by or outsourced to a third party; and proprietary, off-the-shelf solutions. Of the user tools, 50% were developed in-house, 25% were based on open-source solutions and 25% were sourced from proprietary software companies. In 28% of cases, developers used a combination of open-source, proprietary and bespoke tools.

Proprietary tools used by institutions included Zoomify, Tableau, InstantAtlas, 3D Issue, as well as a Flash-based exhibitions outsourced to a third party and tools or modules provided through systems such as eMuseumPlus. Open-source tools included OpenZoom, OpenStreetMap, TimeMap, Sibelius Scorch, Fedora GSearch, Solr and the use of Google Maps to layer historical maps and images. Tools developed in-house included discussion forums, user annotations, geo-mapping and spatial analysis.²⁹ The Irish Qualitative Data Archive's anonymisation tool was developed in-house and is free to use, representing an important development in assisting researchers in preparing social science data for archive and reuse while meeting strict ethical standards.

²⁸ Use of social media is not included in these figures.

²⁹ For example, see AIRO's census mapping tools, available at <http://airomaps.nuim.ie/flexviewer/?config=Census2011.xml> (accessed on 31 August 2012).

Curated exhibitions: contextualising and interpreting material

The tools mentioned above are an important way for users to engage with content and can enhance resource discovery, as well as the user's experience. Another method to achieve this, and to add value to a collection, was through the development of online curated exhibitions of digital assets, which provided users with novel ways of interacting with primary material. This method also provided the host institution with an opportunity to display collections that might otherwise be temporary floor exhibitions or might not be displayed at all. The National Archives of Ireland, the National Library of Ireland and RTÉ were among a growing number of institutions to provide this feature to their users, while a number of interviewees, including the IFI, shared their plans to develop this feature in the future. Malachy Moran from the RTÉ Sound Archive told us that online curated exhibitions are a 'way of giving value [to] the archive without compromising responsibilities to preserve the material' while providing an entertainment element to the archives. He added that designing an online exhibition required considerable resources and a great working knowledge of the collections but felt that the return was worth it, especially as online exhibitions helped to fulfil the archive's remit of providing public access to its holdings.

Curated collections or exhibitions enable content managers to expose the richness of their collections while maintaining control over what content is actually made available. This is a useful method to employ, especially in situations where copyright or access policies, issues and restrictions persist. They act as an advertisement for an institution's holdings and can result in increased footfall, as online exhibitions, which often link multiple types of media objects, reveal the diversity of the content holder's collections. As an example, a podcast of a lecture linked to supporting documents, audio snippets or moving images showcases the archive but also enhances the users'/students' experience and their engagement with knowledge. An Foras Feasa at NUI Maynooth has already successfully developed and enhanced its repository with such features. Damien Gallagher, currently a senior software engineer at the DRI but previously the senior developer for An Foras Feasa's CRADLE (Collaborate, Research, Archive, Discover, Learn, Engage), explained the use of 'RDF [Resource Description Framework] to describe the relationships between each object in [the repository]', in order to generate graphs to illustrate the connections between different types of objects, including documents, audio, moving images,

**Curated collections
or exhibitions
enable content
managers to
expose the richness
of their collections
while maintaining
control over what
content is actually
made available.**

user-generated annotations, digital articles, PDFs and slides. Other tools developed in the current Flash-based system (a move to HTML5 is being considered) include a discussion forum and a slideshow tool. Similarly, the National Library of Ireland produced a Flash-based online exhibition for the 90th anniversary of the 1916 Rising and stated that this was an area that it would like to develop. It viewed online exhibitions and curated collections as an important platform for contextualising and interpreting its material. The Irish Museum of Modern Art has also produced a number of virtual tours that document past exhibitions and collections.

Online curated collections or exhibitions can also support cooperation between different institutions, and one respondent identified this as an activity that the DRI should support.

Going mobile: creating 'mobile-friendly services'

The development of mobile applications was discussed by a number of institutions. Although few had fully developed mobile applications that are available to download and use, many referred to this as a feature for future development, to complement and enhance their digital, as well as physical, assets and holdings. The University of Limerick informed us of its 'Ireland under siege' app, released in May 2012 and developed in conjunction with NUI Galway and the Royal Irish Academy.³⁰ The app, which brings to life the landscape of important battle sites, was developed for Apple iOS as well as Android, and is supported by a website and a blog, 'Immersive learning in history', which documents the 'production of [the] augmented reality mobile phone application'.³¹ The mobile app and the website encourage 'enquiry-based learning in historical research',³² while the blog, which reveals some of the technical choices and features, encourages users and students to engage with software development.

Libraries also described the development of 'mobile-friendly services' and apps to provide access to catalogues, as well as other resources. One librarian informed us that 15–20% of the catalogue's hits were now from mobile devices, and therefore mobile apps and QR codes represented an important way to keep up with user demands and the changes in user activity. Clare County Library also discussed developing 'a version of [its] catalogue for mobile devices' and was considering developing a mobile app for its map collection.

³⁰ See <http://www.irelandundersiege.com> (accessed on 21 August 2012).

³¹ See <http://testarea.edublogs.org/about/> (accessed on 21 August 2012).

³² 'UL launches "Ireland under siege" mobile phone app', available at <http://www.ul.ie/news-centre/news/ul-launches-ireland-under-siege-mobile-phone-app> (accessed on 21 August 2012).

Visualising geospatial data and other mapping tools

The use, creation and visualisation of geospatial data, as well as geo-browsing, were discussed by 16% of interviewees, including Justin Gleeson, Project Manager of AIRO, which gathers and analyses spatial datasets from across Ireland. He informed us of AIRO's developments, which included the use of different mapping and visualisation tools, and stated that it had considered various open-source tools to help with the visualisation and interactivity of datasets but had opted for proprietary software, InstantAtlas and Tableau, as a more cost-effective solution because it provided user support. AIRO, however, also used open-source software such as Drupal and MySQL and developed a Flash-based tool to query datasets. Examples of AIRO's mapping tools include Census Mapping, a Crime Mapping Toolkit and a Social Housing interactive map and analysis tool.³³

The Clare County Library website includes a tool, Clare GMaps, which uses Google Maps to layer historical maps of Clare over current Google images. The site also includes a tool called MapBrowser, which provides users with access to a number of historical maps, including David Rumsey's maps of Clare. The Oral History Network of Ireland also referred to a recent oral history project, which plotted oral accounts of a particular area to a specific location.

Building an online community: social media

Nearly all of the institutions we interviewed used social media to engage and interact with the wider public and their online user base. Social media was used to notify audiences of events, publications and other news, but it also provided institutions with an opportunity to encourage and develop relationships between and among members of their online community.

In a few cases, institutions that used social media often had dedicated staff to manage, populate and control the various sites. The National Library of Ireland informed us that its use of social media sites such as Facebook, Twitter and Flickr was of huge benefit. It not only attracted readers to the library but also helped it to enrich the catalogue's metadata: for example, users of Flickr left comments and in some cases identified unknown places, landmarks and individuals in historical photographs. Users also helped to date certain images and tagged and identified material objects, such as pioneer pins and flags. Without this type of user-generated content, much of the context of these historical photographs would never have been retrieved. This type of user engagement not only enriched the National Library of Ireland's photographic

³³ See <http://www.airo.ie> (accessed on 21 August 2012).

collection but also provided users with a unique opportunity to engage with the library's content and to contribute to and enhance their national cultural (digital) heritage. The National Library of Ireland also informed us that, after the publication of an article in the *Irish Times*, its Flickr Commons' photostream received 40,000 hits in one day alone, demonstrating how social media can facilitate greater user engagement with archival material (in contrast, in 2010, an average of 507 people per day visited the library's premises on Kildare Street).³⁴

Future developments of user tools: fulfilling user expectations

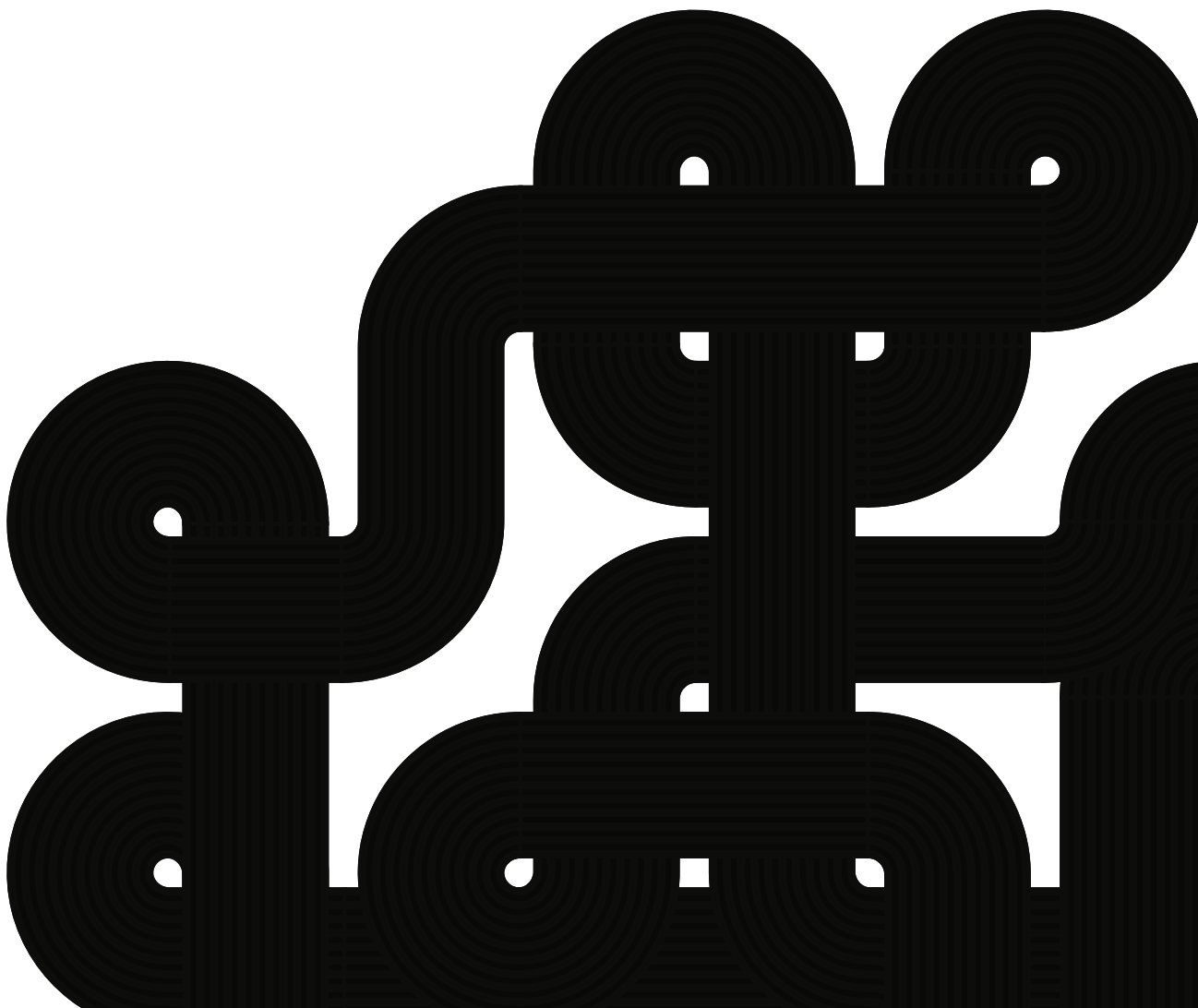
New user tools and resources to engage the public and the younger generation, including primary, secondary and third-level students, as well as academics and scholars, with digital content and enhance user experiences were an area that our interviewees identified as crucial for future development. Particularly, a number of organisations sought to build educational tools including games and other interactive resources. The provision and development of new learning objects and other educational resources was also mentioned. Temporal and spatial tools were also discussed, including the development of time maps and timelines to compare documents, audio etc. over different time periods, as well as geographical mapping of content to map collections and the landscape to allow users to interact with content through a map interface. Mapping tools were also discussed in conjunction with the use of crowd-sourcing to develop collections of special content (e.g., images of special architecture). The development of user workspaces, or 'light boxes', that allow individuals to curate their own private collections or exhibitions was also discussed by some of our interviewees, especially those working with vast photographic or art collections. Resource discoverability was also high on organisations' agendas, and a number of librarians spoke of future developments to optimise search queries and results and to support distance reading. A number of our interviewees wished to develop mobile applications, including apps for smartphones and tablets. One respondent noted that, although this was an area that it wanted to develop, it would look for platform-neutral applications to counter socio-economic divisions. Although our interviewees did not discuss the use of third-party APIs (application programming interfaces) for the creation of user tools, this is an important area of resource development. APIs also enable different systems to share metadata and data as they support the harvesting of content from trusted partners and institutions.

³⁴ 'Irish museums and galleries boast 3 million visitors in 2010', *Journal.ie*, February 2012, [http:// www.thejournal.ie/irish-museums-and-galleries-boast-3-million-visitors-in-2010-90608-Feb2011](http://www.thejournal.ie/irish-museums-and-galleries-boast-3-million-visitors-in-2010-90608-Feb2011) (accessed on 31 August 2012).

In this review, there was a strong sense that the development of user tools was a priority, and as discussed, many interviewees have already embarked on providing their users with innovative and novel ways to access, use and visualise digital content. Others wished to develop and enhance their digital collections with user tools and resources but were unsure of how they should do this. Discoverability and access to resources and objects was identified as

of key importance, and the use and development of different user tools with various functionalities was viewed as one method of achieving this goal. As an 'interactive' digital repository, the DRI will provide a number of specific, targeted user tools and will also support the development of resources through its API. This activity will enhance and enable collaboration between the DRI and the community while enriching the end-user's experience, thus achieving a goal shared by our interviewees.

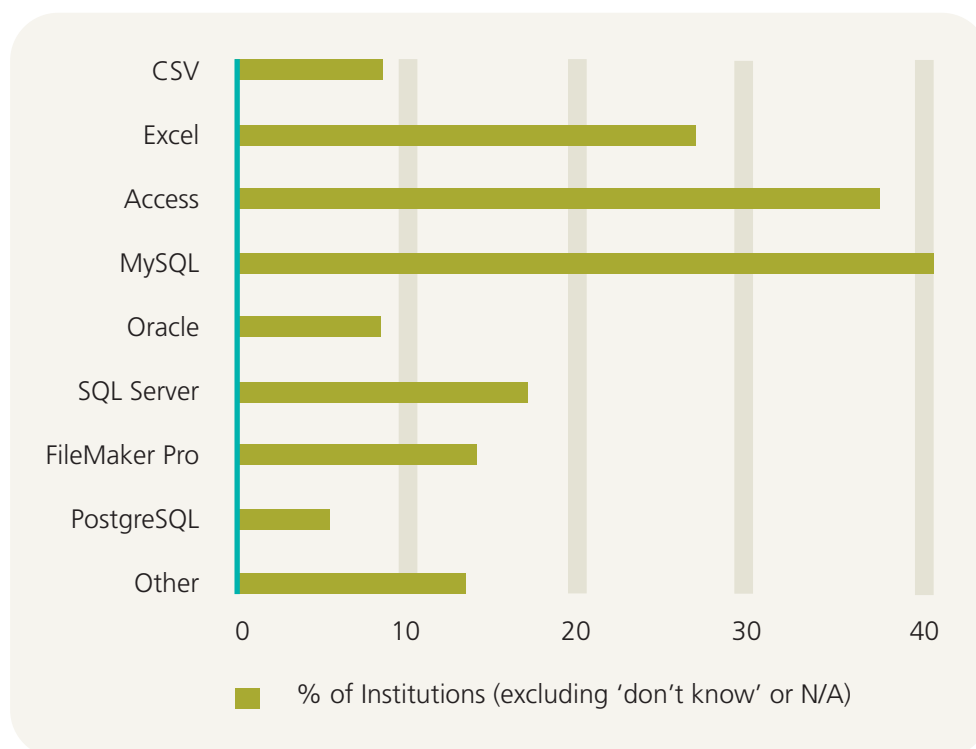
**Discoverability
and access to
resources and
objects was
identified as of
key importance...**



Structuring content: database formats/systems, content management systems and repository software

Figure 7 provides an overview of the database formats and systems used by our interviewees.³⁵ Microsoft Excel and CSV files are deliberately included; although neither is a database system, they are used by institutions for a variety of reasons, e.g., Excel spreadsheets for in-house management, crowd-sourced transcriptions or legacy catalogues, and CSV for storing raw data such as GIS or supporting the import/export of data. The potential use of CSV may be higher, given that it is a common export format. 'Other' database systems include Mulgara, eXist-db, Basis and unspecified GIS database formats.

Fig. 7: Database formats and systems



³⁵ Sources of further information on the database and content management systems mentioned in this section can be found in Appendix 2.

Many of the institutions used more than one database system. Microsoft Access was used by 37% of our interviewees. In most cases, however, Access was a legacy database that was in the process of being upgraded or migrated to a new system. The IAA informed us that Microsoft Access worked well for its in-house catalogue, but its 'biggest bugbear' was that the catalogue 'is not available online' and 'Access is simply not robust enough to put online'. To resolve this problem, the archive is migrating to Adlib, which will enable online access to the catalogue. Adlib, a library management system, is also used by the National Museum of Ireland and the National Archives of Ireland. Another archive also used Microsoft Access for its in-house catalogue, which at present is available only locally in the reading room, but expressed similar aspirations to make the catalogue more widely available.³⁶ This archive's priority in choosing a new system was 'openness...that it would have open database connectivity', that is, it would have the ability to support interoperability and connectivity between systems.³⁷ Some archives had discussed the use of open protocols, particularly OAI-PMH, as a method to share and harvest metadata. The archive was also keen to understand what library management systems other institutions were using or migrating to, stating that it is important to have a 'broad national perspective on things...so if there are a lot of institutions moving...[in the same direction], we would move in a very coherent way'.³⁸

The prevalence of open-source software and solutions is also reflected in Fig. 7, as MySQL is the most common relational database management system (RDMS) in use at the moment. Open-source database solutions not only are affordable but also can have advantages over proprietary software in terms of long-term preservation of and access to database content. The use of RDMSs also reflects international trends and SQL's dominance in the last number of decades. None of our interviewees mentioned NoSQL ('Not only SQL'), however, which is designed to be web-scalable.³⁹

Open-source database solutions not only are affordable but also can have advantages over proprietary software in terms of long-term preservation of and access to database content.

³⁶ Anonymous institution.

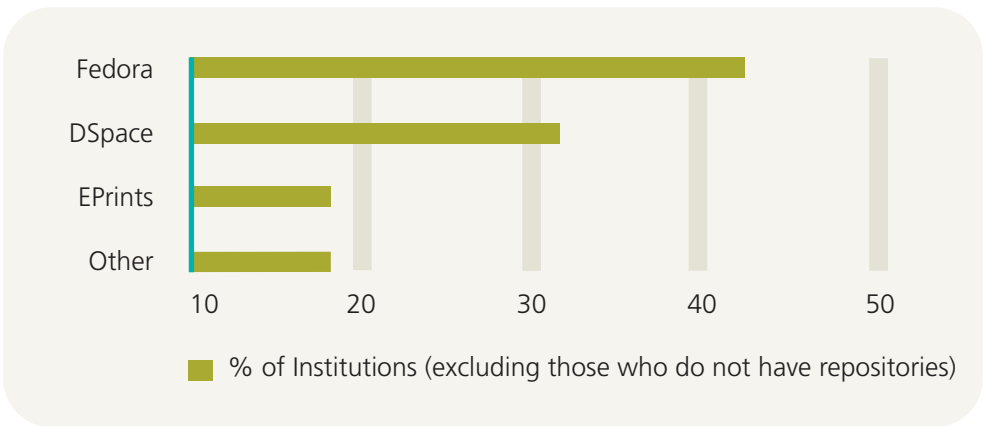
³⁷ *Ibid.*

³⁸ *Ibid.*

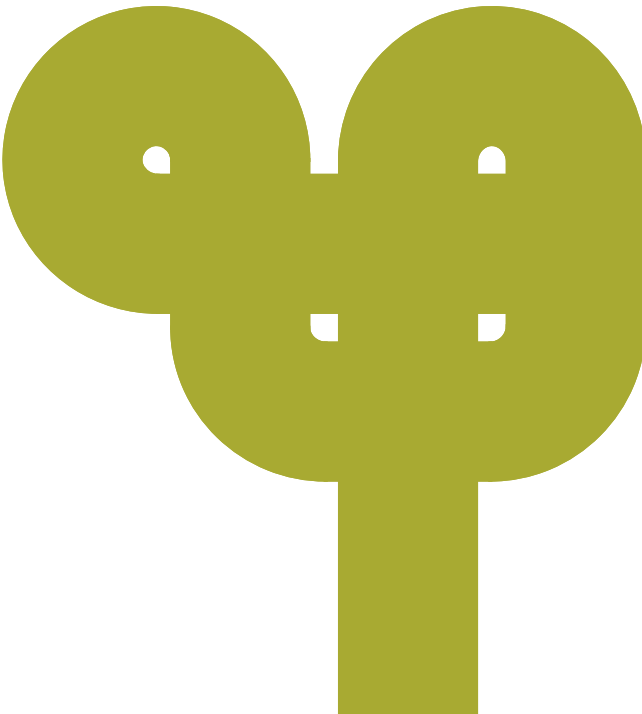
³⁹ Audrey Watters, *The age of exabytes: tools and approaches for managing big data* (2010), available at http://www.readwriteweb.com/archives/download_our_latest_free_report_the_age_of_exabytes.php (accessed on 21 August 2012).

Of the institutions that we have spoken to thus far, 38% had their own digital repository. Included in this number are institutional repositories that deal with academic output, e.g., theses and articles, but repositories also included other datasets, such as those for research-based projects or national cultural institutions. As Fig. 8 shows, the open-source digital asset management system Fedora Commons was the most popular, used by 44%. DSpace was used by 31%, and EPrints by just under 19%, both also open-source. Other repository systems included TYPO3, VTLs VITAL (a Fedora-based system) and other, unspecified commercial solutions.

Fig. 8: Digital asset management systems



Our interviewees used an array of content management and front-end systems, including commercial, open-source and custom in-house solutions such as Drupal, WordPress, Fez, ExpressionEngine, eMuseumPlus, ArtBase, Adlib and Kentico. It is noteworthy that there is no consensus on the content management systems used, reflecting the wide diversity and variety available.



Absences

There were some absences in the interviews (although it should be noted that an absence in the interview does not necessarily reflect an absence in practice). The interviewees did not discuss the preservation of data obtained or modified through the provision of user tools, nor policies that would ensure preservation and future maintenance of those tools. Tools were considered primarily as user tools, and not to facilitate ingest. There was little mention of possible intermediary machine formats or middleware that might be required by a front-facing application or of a desire for standard APIs.

There was no mention of the processes by which policies were developed and implemented, and few⁴⁰ mentioned delivering in-house education or training programmes aimed at raising staff skill levels. Linked data, an emerging practice to enable the web to connect related data that were not previously linked, was rarely mentioned.⁴¹ It is difficult to draw conclusions from these absences; although they could reflect absences in the institutions, they could also reflect a methodological limitation (put simply, not all of the questions that we asked could be answered by the person whom we interviewed). The absences are noted here, as the DRI will need to explore these issues in more depth in its future work.

⁴⁰ The Digital Humanities Observatory was a notable exception.

⁴¹ An exception is RTÉ, which is engaged in a linked data project proposal with the DRI and the Digital Enterprise Research Institute. In stakeholder consultation, Marie Wallace, Social Analytics Strategist at IBM, emphasised the need to consider a linked approach to sharing data.

Conclusion

A variety of institutions in Ireland are tasked with caring for digital content. This content is varied in nature; much is rare, unique and valuable. In the face of a rapidly changing technological field and considerable resource limitations, organisations are cognisant of the necessity to develop robust workflows in order to protect their digital collections and ensure that their richness is available to future generations. Many organisations are eager to add value to the resources in their care. Digital archives are moving beyond providing simple 'access' towards identifying ways in which they might transform their content to meet changing user needs.

Many are aware of the challenges that they face. Digitisation and digital projects are often based on short-term funding, with few resources available to ensure long-term sustainability of these projects. Skills deficits were evident, particularly in specialised technical areas. The digital field also changes rapidly, which creates difficulties in maintaining technical infrastructures over time.

A key problem highlighted by many of our interviewees was the difficulty in preserving born-digital content. Digital storage systems are not particularly robust and degrade over time or become unusable as the technology evolves. Digital formats change and become inaccessible; in some cases, archival formats were not available. In many instances, particularly with respect to e-mail, it was not clear what digital content should be saved for future generations and what could be safely discarded. In the absence of clear guidelines, ad hoc and sometimes erratic decisions were being made.

There were concerns that it might be assumed that digitisation would replace or supersede the conservation of the physical objects, which should not be the case. Digital objects are harder to preserve over time than physical objects. The value of digital objects comes instead from their ability to be found and shared across collections and manipulated and interrogated by users. This value is severely limited, however, if collectors do


Digital objects are harder to preserve over time than physical objects. The value of digital objects comes instead from their ability to be found and shared across collections and manipulated and interrogated by users.

not apply the same metadata standards or if the data are contained in very different formats. This survey indicated that ensuring interoperability is a critical challenge for the Digital Repository of Ireland and its stakeholder community. The content that is in most danger is born-digital, content that is being created right now but is under threat of being lost to future generations (the 'digital black hole'). It is highly likely that some born-digital content is already lost and may be unretrievable.

A number of tensions were also evident. In the context of insufficient staffing and IT support, digitisation programmes were driven by the immediate value returned for funding received, with the selection of collections often on the basis of high profile rather than a systematic approach to meet the wider community's longer-term needs. Whole-scale digitisation of collections was not being considered; if anything, digitisation projects were being scaled down or placed on hold. There is a further resource-driven tension between imaging projects and digitising projects. Imaging creates digital images of objects; it does not include the extra steps required by digitisation, such as the addition of metadata, transcriptions and contextual data and the secure storage of the objects created. The decision to image has in some cases led to the need to catalogue objects retrospectively to enable their reuse.

In developing content management systems, although there was a desire for more online content, there was also a desire to maintain control over content use and reuse in many cases. If only in-house access to databases is offered, it is much easier to control how data can be used. Inevitably, putting data online entails a certain loss of control. Many institutions wished to enable open access to their data, yet some of their content had ethical or copyright restrictions attached to it that made open access impossible. Finally, although most interviewees spoke of their willingness to share either their metadata or their data, this willingness was matched by the necessity of maintaining a link to the data, as retaining ownership was important in terms of creating their own brand and obtaining further funding for their institution. The nature of this link, whether it is attribution, citation and/or exposure to analytical software, will need to be explored in further discussions. While maintaining their own identity and brand, many institutions were also participating in large-scale international aggregated digital archives and catalogues, such as Europeana,⁴² recognising the value of an international profile. In addition, a significant trend internationally towards sharing publicly generated data is evident, and as new copyright and ethical frameworks are developed, barriers that militate against sharing may lessen.

⁴² See <http://www.europeana.eu/portal/aboutus.html> (accessed on 21 August 2012).



A number of trends were evident. Many of the interviewees were increasingly considering open-source solutions to their data management problems, in part because open-source is more affordable but also to avoid future archival problems attached to using proprietary systems. Many of the interviewees had legacy systems in place yet wished to adopt new technology, particularly in terms of enabling greater use of the internet. A key challenge faced by many was how to go about integrating multiple different systems (for example, different catalogue record systems).

Many of the interviewees were forward-looking in seeking new ways to develop and enhance the resources under their care. Many were interested in developing online educational tools, with a few engaged in or expressing interest in creating collaborative digitisation programmes, such as crowd-sourced annotation, metadata generation, and transcription or imaging of content. A number of institutions were using social media (particularly Twitter and Facebook) to generate audience awareness and interest in their collections.

The Irish digital landscape is rich, and the range of data held is diverse and impressive. It encompasses music, films, radio and television programmes, manuscripts, maps, art, architectural drawings, correspondence, interviews, archaeological surveys, newspapers, diaries, images (including 3D) and new media, including user-generated content. In the absence of a national strategy for protecting our digital cultural and social heritage, these objects are in danger of being lost to future generations. This is the DRI's contention, and our goal for 2013 and 2014 is to work with the community in the development of national guidelines for digital preservation and access, in order to inform future policy for our cultural and social digital heritage.

In the absence of a national strategy for protecting our digital cultural and social heritage, these objects are in danger of being lost to future generations.

Appendix 1: Methodology

Stakeholder interview sample selection

An intrinsic part of information-gathering for both requirements and policy is conducting stakeholder interviews. By considering the policies and practices already in place, the DRI can develop in a way that is supportive of existing institutions and projects. The DRI potentially includes data types and collections from various cultural, national and independent institutions that all have very special (although some overlapping) requirements and policy concerns for data ingestion, functionality, preservation and access. Other content providers include private or individual researchers. In addition, the DRI must satisfy those who wish to use the system and its collections.

Organisations interviewed to date (2012)

All-Island Research Observatory, National University of Ireland Maynooth

An Foras Feasa, National University of Ireland Maynooth

Clare County Library

Crawford Gallery, Cork

Digital Humanities Observatory, Royal Irish Academy

Discovery Programme

Dublin City Archives

Dublin City University Library

Economic and Social Affairs Institute

Health Research Board

Hunt Museum, Limerick

Irish Architectural Archive

Irish Film Institute

Irish Manuscripts Commission

Irish Museum of Modern Art

Irish Qualitative Data Archive, National University of Ireland Maynooth

Irish Qualitative Data Archive/National Institute for Regional and Spatial Analysis, DRI
Demonstrator Project, National University of Ireland Maynooth
Irish Traditional Music Archive
National Archives of Ireland
National Centre for Technology in Education
National Folklore Collection, University College Dublin
National Gallery of Ireland
National Irish Visual Arts Library, National College of Art and Design
National Library of Ireland
National Museum of Ireland
National University of Ireland, Galway, DRI Demonstrator Project
National University of Ireland, Galway, Library
National University of Ireland Maynooth, DRI Demonstrator Project
National University of Ireland Maynooth, Library
Oral History Network of Ireland
Raidió na Gaeltachta
Royal Irish Academy
RTÉ Digital
RTÉ Sound Archive
School of History and Archives, University College Dublin
School of Information and Library Studies, and Earth Institute, University College Dublin
Trinity College Dublin, Library
University College Dublin, Library
University of Limerick, Library and Department of History

Ethics approval/consent form for user/stakeholder interview recordings and transcriptions




Respondent information sheet (in-depth qualitative interviews)

Digital Repository of Ireland (DRI): key consultations

Thank you for agreeing to participate in this study. The DRI research consortium is tasked with building a robust, scalable, accessible and sustainable trusted digital repository (TDR) and access repository for the humanities and qualitative social sciences. As part of our work towards meeting this goal, we are documenting the experiences, concerns and preferences of key stakeholders in the research, library and archiving process. Your contribution in this regard will be extremely valuable.

The research is being carried out at the National University of Ireland Maynooth.

The investigators are:

-  Dr Aileen O'Carroll, Policy Manager, DRI/National University of Ireland Maynooth
-  Dr Sharon Webb, Requirements Manager, DRI/National University of Ireland Maynooth
-  Dr Sandra Collins, Director, DRI/Royal Irish Academy

With your permission, the interview will be recorded. Afterwards it will be written up/transcribed. Both the recording and the interview notes/transcription will be stored in a locked cabinet in the project head office at NUI Maynooth.

Once all the interviews are completed, the audio recordings and interview notes/transcripts will be deposited in an archive, where other bona fide researchers may consult them. You may be happy to be personally identified in these public materials. However, if you wish, your name will be removed, and your comments made unattributable.




Once again, we thank you for your participation. However, it is important for you to know that your participation in the research is entirely voluntary. You may withdraw your consent to participate at any time, without obligation.

Having read this information sheet, please read and sign the consent form.

Consent form (in-depth qualitative interviews)

Project Title: Digital Repository of Ireland (DRI): key consultations

The investigators are:

-  Dr Aileen O'Carroll
-  Dr Sharon Webb
-  Dr Sandra Collins

Material gathered during this research will be treated as confidential and securely stored in a locked cabinet at NUI Maynooth. You have the right to access any of your interview materials (tapes, transcripts and notes) at any time.

Please answer each statement below concerning the collection of the research data.

1.	I have read and understood the information sheet.	Yes 	<input type="checkbox"/>
		No 	<input type="checkbox"/>
2.	I have been given the opportunity to ask questions about the study.	Yes 	<input type="checkbox"/>
		No 	<input type="checkbox"/>
3.	I have had my questions answered satisfactorily.	Yes 	<input type="checkbox"/>
		No 	<input type="checkbox"/>
4.	I understand that I can withdraw from the study at any time without having to give an explanation.	Yes 	<input type="checkbox"/>
		No 	<input type="checkbox"/>
5.	I agree to the interview being audiotaped and to its contents being used for research purposes.	Yes 	<input type="checkbox"/>
		No 	<input type="checkbox"/>

Below are sets of statements that give you, the interviewee, a series of options about how you wish your interview to be used. Please answer each statement.

6. I agree to being identified in this interview and in any subsequent publications or use.

Yes

☐

No

☐

If you answered 'Yes' to Q. 6, please go directly to Q.8

If you answered 'No' to Q. 6, please also answer Q.7

7. Where used, my name must be removed and my comments made unattributable.

Yes

☐

No

☐

8. I agree to the interview notes/transcripts (in line with the conditions outlined above) being archived and used by other bona fide researchers.

Yes

☐

No

☐

9. I agree to my audiotapes (in line with the conditions outlined above) being archived and used by other bona fide researchers.

Yes

☐

No

☐

10. I would like my name acknowledged in the report and on the project website (without linking it to content or quotation).

Yes

☐

No

☐

Name (printed) _____

Signature _____ Date _____

Your contribution is greatly appreciated. Feel free to contact us if you have any further questions.

Dr Aileen O'Carroll: Phone: (01) 708 3596/E-mail: aileen.ocarroll@nuim.ie

Dr Sharon Webb: Phone: (01) 708 7182/E-mail: sharon.webb@nuim.ie

Dr Sandra Collins: Phone: (01) 609 0668/E-mail: s.collins@ria.ie






If during your participation in this study you feel the information and guidelines that you were given have been neglected or disregarded in any way, or if you are unhappy about the process, please contact the Secretary of the National University of Ireland Maynooth, Ethics Committee at pgdean@nuim.ie or (01) 708 6018. Please be assured that your concerns will be dealt with in a sensitive manner.

Topic guide for user/stakeholder interviews

The key approach is to use open-ended questions (e.g., can you tell me about/can you describe?), following the flow of the interviewee and only directing if the issues that need to be discussed do not emerge naturally in the course of the conversation.

We consider the resource/archive in terms of its current data lifecycle. Our aim is to establish how users/stakeholders currently support their digital resources/objects and how they develop and maintain their data archives/repositories. This will assist the DRI in setting key objectives and priorities.

STAGES:

- Pre-ingest:  The activities surrounding the data before they are prepared for archiving.
- Ingest:  Preparation and deposit of data into archive.
- Preservation:  Fulfilling the archive's responsibility to preserve data.
- Dissemination/Reuse:  Fulfilling the archive's responsibility to enable dissemination/reuse of data.
- DRI:  Future development in a federated repository.

KEY TOPICS

QUESTIONS

Stage in archive lifecycle: Pre-ingest

Digital objects/resources. Quantity, data formats (.txt, .doc), processes of digitisation (crowd-sourcing)

Computer or software systems in use

User interfaces (bespoke, particular product)

Static or living archive

Bilingual data

Can you tell me about your resource/archive/repository?

Can you describe your data/content?

Are all your data digitised?

Can you describe the digitising process?

Can you describe the current system you use for your data collection?

How do you envisage your resource developing in the future?

Expected data growth (in a two-year period, for example)

How much data are there now for ingest into the DRI?

Data quality assessment/quality control process (in terms of data formats and data content)

How do you assess data/content quality?

Stage in archive lifecycle: Ingest

Nature of data (special concerns, sensitive data, rarity, commercial issues). Access issues/policy

In terms of archiving or storing your data, are there any particular concerns or considerations? How did you address them?

KEY TOPICS

QUESTIONS

Ownership/copyright

Who owns the data? Are there copyright issues? Do you have licensing agreements?

Intellectual property

Are there any IP issues?

Collection priorities

How do you source the data?

Do you have special priorities?

Catalogue ontology/thesaurus

Have you developed a catalogue? If so, can you describe it?

Metadata formats

What metadata standards do you use?

Database formats

Do you know what database system you are using (MySQL, Excel, XML etc.)?

Linked data

Open data

Stage in archive lifecycle: Preservation

Future-proofing; data formats/longevity of data

Can you describe your preservation process, if any?

Data security (physical threats, virtual threats)/redundancy

Where are the data physically stored?

What security systems do you have in place, if any?

What level of redundancy (people, software, hardware, organisations) would you like to see implemented so that you feel the DRI is trusted, e.g., is two copies of data safe, three copies?

Do we need more than one person who is able to develop/maintain the system?

KEY TOPICS

QUESTIONS

Budgets

Do you include budget lines for preservation/ingest/storing data in a repository in your proposals?

Are you thinking about it?

Stage in archive lifecycle: Dissemination/Reuse

User experience/expectations (students, researchers etc.)

Can you describe who uses your data?

How do you see users in the future?

User tools

Do you provide any tools to enable the user to interact with the data?

Address concerns surrounding data security

How is the DRI 'trusted'?

DRI/organisation and infrastructure

Expectations of the DRI

What is important to you in terms of developing your resource?

Scale of investment to date in the project and predicted future investment

What are your biggest challenges?

Appendix 2: Resources

Formats⁴³

Name	Resource
Formats: general	http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf ; http://www.nationalarchives.gov.uk/documents/selecting-storage-media.pdf
AIFF (Audio Interchange File Format)	http://www.digitalpreservation.gov/formats/fdd/fdd000005.shtml
CAD (Computer Aided Design)	http://www.trixsystems.com/cad.html
DCP (Digital Cinema Package)	http://indieranch.com/Post/DCP_master.html
DGN (Digital Negative Format)	http://www.fileinfo.com/extension/dgn
DjVu	http://djvu.org/resources/
FLAC (Free Lossless Audio Codec)	http://flac.sourceforge.net/faq.html
JPEG (from 'Joint Photographic Experts Group')	http://www.jpeg.org/faq.phtml
JPEG 2000	http://www.jpeg.org/faq.phtml?action=show_answer&question_id=q3d5bc0701c9b6
Microsoft Office	http://office.microsoft.com/en-ie/support/?CTT=97
Microsoft Word	http://office.microsoft.com/en-us/word-help/
MP3	http://telos-systems.com/techtalk/hosted/Brandenburg_mp3_aac.pdf
MPEG (from 'Moving Pictures Experts Group')	http://telos-systems.com/techtalk/hosted/Brandenburg_mp3_aac.pdf
MPEG-2, MPEG-3, MPEG-4	http://telos-systems.com/techtalk/hosted/Brandenburg_mp3_aac.pdf
Open Office	http://incubator.apache.org/openofficeorg/index.html
PDF (Portable Document Format)	http://www.adobe.com/products/acrobat/adobepdf.html
PDF/A (Portable Document Format, Archive Standard)	http://www.adobe.com/enterprise/standards/pdfa/
QuickTime	http://www.apple.com/quicktime/what-is/
RAW	http://www.rondayphotography.com/Understanding%20the%20RAW%20File%20Format.htm
RTF (Rich Text Format)	http://office.microsoft.com/en-us/word-help/about-rich-text-format-documents-HP001004477.aspx

⁴³ We would like to thank DRI student interns Sam McGrath and Donal Fallon for their assistance in preparing this appendix.

Name	Resource
SVG (Scalable Vector Graphics)	http://www.w3.org/Graphics/SVG/
TEI (Text Encoding Initiative)	http://www.tei-c.org/index.xml
TIFF (Tagged Image File Format)	http://www.awaresystems.be/imaging/tiff/faq.html
WAV (Waveform Audio File)	http://www.digitalpreservation.gov/formats/fdd/fdd000001.shtml
WCS (Web Coverage Service)	http://www.opengeospatial.org/standards/wcs/
WFS (Web Feature Service)	http://www.ogcnetwork.net/wfstutorial
WMS (Web Mapping Service)	http://www.opengeospatial.org/standards/wms/
WordPerfect	http://www.corel.com/corel/index.jsp
WordStar	http://www.wordstar.org/index.php/wordstar-history
XML (Extensible Markup Language)	http://www.w3.org/XML/

Databases and spreadsheets

Basis	http://www.basis.com/database-management
CSV	http://www.imf.org/external/help/csv.htm
eXist-db	http://exist-db.org/exist/credits.xml
FileMaker Pro	http://www.filemaker.com/company/
Microsoft Access	http://office.microsoft.com/en-us/access/what-is-microsoft-access-database-software-and-applications-FX102473444.aspx
Microsoft Excel	http://spreadsheets.about.com/od/tipsandfaqs/f/excel_use.htm
MySQL	http://www.mysql.com/industry/faq/
Oracle	http://www.orafaq.com/wiki/Oracle_database_FAQ
PostgreSQL	http://www.postgresql.org/docs/faq/
RDMS	http://searchsqlserver.techtarget.com/definition/relational-database-management-system
SQL Server	http://www.microsoft.com/sqlserver/en/us/product-info.aspx

Digital asset/library/content management systems

Adlib	http://www.adlibsoft.com/support/faqs
ArtBase	http://www.artbaseinc.com/faq.swf
Drupal	http://www.3dissue.com/forums/forum/3d-issue-knowledge-base/

Name	Resource
DSpace	http://libraries.mit.edu/dspace-mit/about/faq.html#what
EPrints	http://www.eprints.org/openaccess/
eMuseumPlus	http://www.zetcom.com/products/emuseumplus/?no_cache=1&sword_list[0]=museum
ExpressionEngine	http://expressionengine.com/overview
Fedora Commons	http://www.fedora-commons.org/about
Fez	http://fez.library.uq.edu.au/wiki/Main_Page
FileMaker Pro	http://www.filemaker.com/company/
Kentico	http://www.kentico.com/Company
MySQL	http://www.mysql.com/industry/faq/
Oracle	http://www.orafaq.com/wiki/Oracle_database_FAQ
PostgreSQL	http://www.postgresql.org/docs/faq/
RDMS	http://searchsqlserver.techtarget.com/definition/relational-database-management-system
SQL Server	http://www.microsoft.com/sqlserver/en/us/product-info.aspx
TYPO3	http://typo3.org/about/
VTLS VITAL	http://vitalusers.wikidot.com/
WordPress	http://wordpress.org/about/

Metadata and vocabularies

AARC2	http://www.aacr2.org/about.html
Art & Architecture (Getty)	http://www.getty.edu/research/tools/vocabularies/aat/about.html
DDI	http://libraries.mit.edu/guides/subjects/data/archiving/ddi.html
Dublin Core	http://dublincore.org/specifications/
Dewey Decimal Classification	http://www.oclc.org/dewey/about/default.htm
EBU Core	http://tech.ebu.ch/lang/en/MetadataEbuCore
ESE	http://www.europeana.eu/schemas/ese/
INSPIRE	http://www.esri.com/software/arcgis/arcgis-for-inspire/common-questions
IEEE LOM	http://ltsc.ieee.org/wg12/ ; http://www.dcc.ac.uk/resources/curation-reference-manual/completed-chapters/learning-object-metadata
IPTC	http://www.iptc.org/cms/site/index.html?channel=CH0099
ISAD(G)	http://archiveshub.ac.uk/isadg/

Name	Resource
MARC 21	http://www.loc.gov/marc/faq.html#definition
METS	http://www.loc.gov/standards/mets/METSOverview.v2.html
MODS	http://www.loc.gov/standards/mods/mods-overview.html
NISO MIX	http://www.loc.gov/standards/mix/
RDF	http://www.w3.org/RDF/
SPECTRUM	http://www.pro.rcipchin.gc.ca/GetForumRecord.do?type=sd&lang=en&id=FORUM_23068&ens=cnRsTGFuZz1lbiZydGxUeXBIPXNk
VRA	http://www.vraweb.org/about/index.html

User tools

3D Issue	http://www.3dissue.com/forums/forum/3d-issue-knowledge-base/
eMuseumPlus	http://www.zetcom.com/products/emuseumplus/
Facebook	https://www.facebook.com/help/?page=160699464027593&ref=hcsubnav
Flickr	http://www.flickr.com/about/
Google Maps	http://www.makeuseof.com/tag/technology-explained-google-maps-work/
InstantAtlas	http://www.instantatlas.com/support.xhtml
OpenStreetMap	http://wiki.openstreetmap.org/wiki/Main_Page
OpenZoom	http://www.openzoom.org/ ('No longer maintained or supported')
Sibelius Scorch	http://www.sibelius.com/products/scorch/index.html
Solr	http://lucene.apache.org/solr/
Tableau	http://www.tableausoftware.com/about
TimeMap	http://www.timemap.net/index.php
Twitter	https://twitter.com/about
Zoomify	http://www.zoomify.com/about.htm

All accessed August 2012

Appendix 3: Stakeholder Advisory Group members

The DRI is honoured to have such an expert group to whom to present our progress. The members of this group are as follows:

Name	Role/institution
Mr Nicholas Carolan	Director, Irish Traditional Music Archive
Dr Mary Clark	City Archivist, Dublin City Archives
Ms Patricia Clarke	Senior Policy Analyst, Health Research Board
Ms Fionnuala Croke	Director, Chester Beatty Library
Ms Catriona Crowe	Head of Special Projects, National Archives of Ireland
Mr John Fitzgerald	Librarian and Director of Information Services, University College Cork Library
Mr Chris Flynn	Principal Officer, Cultural Policy, Department of Arts, Heritage and the Gaeltacht
Mr Jonathan Grimes	Information and Digital Services Manager, Contemporary Music Centre
Dr Cathy Hayes	Administrator, Irish Manuscripts Commission
Dr John Howard	University Librarian, University College Dublin Library
Mr Olivier Kazmierczak	ICT Manager, National Museum of Ireland
Ms Beatrice Kelly	Head of Policy & Research, The Heritage Council
Ms Christina Kennedy	Senior Curator, Head of Collection, Irish Museum of Modern Art
Ms Múirne Laffan	Managing Director, RTÉ Digital
Dr Kevin Marshall	Head of Education, Microsoft Ireland
Dr Jason McElligott	Keeper, Marsh Library
Ms Kasandra O'Connell	Head of the Irish Film Archive, Irish Film Institute
Dr Catherine O'Connor	Director and Founding Member, Oral History Network of Ireland
Ms Gobnait O'Riordan	Director, Glucksman Library, University of Limerick
Mr Colum O'Riordan	Archive Administrator, Irish Architectural Archive
Mr Seán Rainbird/ Ms Andrea Lydon	Director, National Gallery of Ireland/Head of Library and Archives, National Gallery of Ireland
Ms Fiona Ross	Director, National Library of Ireland
Mr Paul Sheehan	Director, Library Services, Dublin City University Library
Ms Marie Wallace	Social Analytics Strategist, IBM
Dr Manus Ward	Scientific Programme Manager, Science Foundation Ireland