



## Preservation as a Service

PASIG 2014

Dr Sharon Webb

Knowledge Transfer Manager, Digital Repository of Ireland  
Royal Irish Academy



An Roinn Post, Fiontar agus Nuálaíochta  
Department of Jobs, Enterprise and Innovation



Ireland's EU Structural Funds  
Programmes 2007 - 2013

Co-funded by the Irish Government  
and the European Union

**HEA**

Higher Education Authority  
An tÚdarás um Ard-Oideachas



EUROPEAN REGIONAL  
DEVELOPMENT FUND



## Mission Statement

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



## Vision

DRI links and preserves the rich data held by Irish institutions, providing a central internet access point and multimedia tools



## Funding

Exchequer funded; HEA PRTL1 5, €5.2M

RIA (lead), NUIM, TCD, DIT, NUIG, NCAD

Partners: academic, cultural, social, industry

Sep 2011 – Sep 2015



## Current Status

Preview event held in May 2014 - Staggered release, previews throughout the 2014/2015

UAT being rolled out (phased with releases)

Leveraged projects (ongoing work)

Education and Outreach - (Designated) Community and user engagement

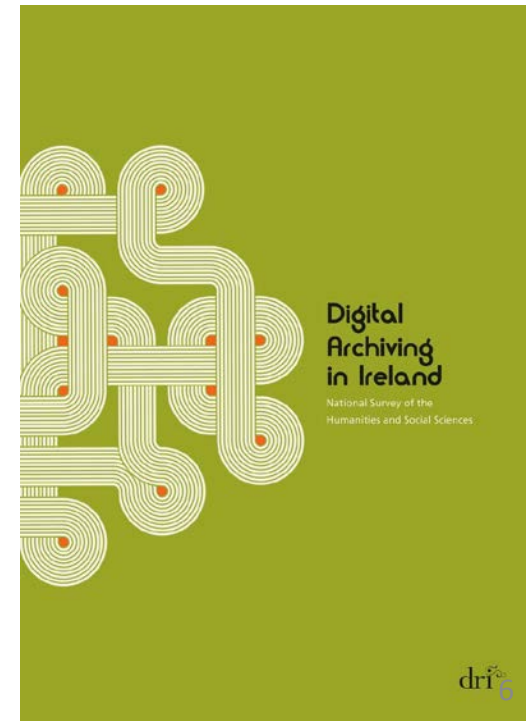
## Various Reports...(core)

*Digital Archiving in Ireland*

*National Survey of the Humanities and Social Sciences*

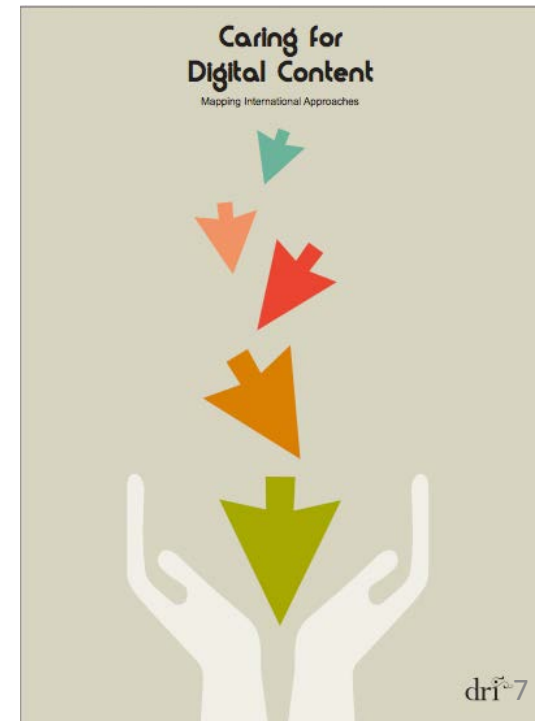
[www.dri.ie/publications](http://www.dri.ie/publications)

Will also be available through repository site.



## Various Reports...(core)

*Caring for Digital Content:  
Mapping International Best Practices*





# And fact sheets...

# Metadata, copyright, file formats, preservation...more to come (including linked data)

October  
2013

2

3

4

## fact sheet

### Metadata and the DRI

DRI uses metadata (often described as 'data about data') to enhance content searchability, discoverability and support overall data management. DRI conducted a comprehensive survey of Irish social science and humanities institutions, and determined the most prevalent metadata standards in use nationally.<sup>1</sup> These metadata standards were mapped to track commonalities, to create guidelines and to ensure DRI's search results are comprehensive and accurate.

**Metadata Standards**  
DRI believes it is good practice to support selected existing metadata standards rather than creating a new metadata standard for the Repository.

DRI will capture, create and utilise descriptive, administrative, structural, technical and preservation metadata.

Phase 1 of our metadata review determined how the Repository will support the following standards: Dublin Core (Simple and Qualified), MODS, EAD, MARCXML.

Phase 2 of this process will work towards the inclusion of additional standards.

**Metadata in DRI**  
DRI will require a number of mandatory fields for all ingested objects to facilitate information retrieval and promote content discoverability. We encourage the creation of rich metadata, and will therefore display all the metadata elements provided for an object or collection.

DRI recommends the use of controlled vocabularies, and has collated a list of those most widely used here: [www.dri.ie/vocabularies](http://www.dri.ie/vocabularies)

**Public Access to Metadata**  
DRI supports the sharing and reuse of data where possible, and all metadata in the Repository will be publicly accessible and licensed under a Creative Commons Attribution Licence (CC-BY).<sup>2</sup> The CC-BY licence is in line with DRI's advocacy for the principles of Open Access.<sup>3</sup>

## fact sheet

### Copyright, Licensing and Open Access

**Copyright** is an area of intellectual Property law which covers original creative works including literary, dramatic, musical and artistic works, film, sound recordings, broadcasts and the typographical arrangement of published editions, computer software and non-original databases, and performances.

Copyright exists from the moment the work is created, and does not require any registration. Generally, copyright covers a work until 70 years after the death of its creator. For digital objects, there may be separate copyrights covering the original work, the new digitised copy, and any associated metadata.

Licensing allows copyright owners to permit approved use and reuse of their work, without waiving their copyrights fully. Licensing can permit both commercial and non-commercial reuse of a work, depending on the terms of the licence, and licences may last in perpetuity or for a specified period, or cover certain geographic locations.

**Creative Commons** is a charitable organisation which provides a suite of free, standardised licences which allow content creators to share their work. Currently there are 6 licences available, which give varying degrees of re-use permissions (see chart on page 2).

The most open licence is CC-BY (Attribution) which allows any re-use of the work as long as the original creator is credited. The most restrictive licence is CC-BY-NC-ND (Attribution-Non-Commercial-NoDerivs) which requests that the creator of the work is credited, and does not allow any derivatives or commercial re-use of the work.<sup>4</sup>

DRI supports the application of standardised licences such as Creative Commons to digital objects, publications and metadata where possible.

**Public Domain Dedication, CC0, or "No Rights Reserved"** is similar to Creative Commons licences, but it does not reserve any rights in a copyright work. Assigning a Public Domain Dedication to a work waives all rights

DRI Factsheet Series No. 1, Metadata and the DRI, Oct. 2013

For more information, please contact:

The Digital Repository of Ireland  
Royal Irish Academy  
19 Dawson St,  
Dublin 2

web: [www.dri.ie](http://www.dri.ie)  
email: [dri@ria.ie](mailto:dri@ria.ie)  
twitter: [@dri\\_ireland](https://twitter.com/dri_ireland)

DRI Factsheet Series No. 2, Copyright, Licensing and Open Access, Feb. 2014

For more information, please contact:

The Digital Repository of Ireland  
Royal Irish Academy  
19 Dawson St,  
Dublin 2

web: [www.dri.ie](http://www.dri.ie)  
email: [dri@ria.ie](mailto:dri@ria.ie)

<sup>1</sup> <http://www.dcc.ac.uk/resources>

## fact sheet

### DRI File Formats

**DRI and Digital File Format Choices:**

As part of DRI's overall strategy to be a certified Trusted Digital Repository (TDR), DRI is developing policy guidelines to assist depositing institutions in their long term digital preservation. These include recommendations regarding the file formats that will be accepted for ingestion into DRI.

File format choices are an extremely important component of any digital preservation strategy. As technology develops, digital formats can change quickly and old formats become unreadable. If they are proprietary formats, long-term support for them may not be guaranteed as companies disappear or simply stop supporting older versions of a particular format. For this reason, DRI's strategy is to recommend formats that are as sustainable as possible in the long term.

**File Formats in DRI**  
The DRI policy for ingestion of file formats will be reviewed regularly to reflect changes in technology and data collections. Below is a list of the file formats currently accepted for

DRI Factsheet Series No. 3, DRI File Formats, March 2014

For more information, please contact:

The Digital Repository of Ireland  
Royal Irish Academy  
19 Dawson St,  
Dublin 2

web: [www.dri.ie](http://www.dri.ie)  
email: [dri@ria.ie](mailto:dri@ria.ie)  
twitter: [@dri\\_ireland](https://twitter.com/dri_ireland)

ingestion into DRI. Preferred formats are recommended archival formats, best suited for long-term digital preservation; accepted formats are commonly used, but not ideal for long-term digital archiving. They are often lower quality or proprietary. If you have a choice, we recommend capturing in preferred formats.

DRI does not recommend converting files from an accepted to a preferred format (for instance, from MP3 to WAV, or JPEG to TIFF), as this will not improve quality, and may in fact reduce it.

**Textual:**  
Preferred: PDF, rtf, txt, xml  
Accepted: PDF, DOC, DOCx

**Audio:**  
Preferred: WAV, BWAV  
Accepted: MP3 MP2

**Image:**  
Preferred: TIFF  
Accepted: JPEG

It is important to stress that these are the formats DRI is currently recommending, and the range of ingestible formats will continue to grow over time.



## fact sheet

### Long-term Digital Preservation

**DRI is a trusted digital repository (TDR)**

As a TDR, we are acutely aware of the problems and risks associated with long-term digital preservation and the activities required to provide sustained access to digital objects.

**Digital Fragility & Dependency**  
The fundamental problem associated with long-term digital preservation is the volatile nature of the digital medium and the inherent reliance on peripheral resources (e.g. software and hardware) to access and display digital objects. Digital objects are fragile and open to corruption.

**Hardware and Software**

Most analogue objects, such as printed ephemera, documents, books, etc., can be used and referenced in and of themselves. Digital objects, however, are software and hardware dependent and therefore only function within certain environments. For example, a printed book can be read without the

printing press - it is functionally independent. An ebook, however, can only be read within a certain content - on a particular device, through a particular programme.

When we preserve digital objects we are essentially preserving access to, and the functionality of, that object. A digital file may exist on a storage medium but that does not guarantee that we can access the contents of that file. Therefore, additional back ups or replication of a digital object does not constitute digital preservation - although it is part of the overall process.

**Volatility and Degradation**

The volatile nature of storage media (magnetic or optical) and the degradation of software (if it is not maintained) contributes to the corruption of the data or the process known as 'bit rot' or 'bit flip'. This corruption can render a digital object inaccessible. It is in this respect that digital preservation cannot be a retrospective activity. Therefore, when we speak of long-term

For more information, please contact:

The Digital Repository of Ireland  
Royal Irish Academy  
19 Dawson St,  
Dublin 2

web: [www.dri.ie](http://www.dri.ie)  
email: [dri@ria.ie](mailto:dri@ria.ie)





Our “services” are dictated by a number of crucial elements

1. Community needs (member and/or (end) user?)
2. Sustainability - as a four year funded ‘project’
3. External factors - open data, open government, H2020 (and other funding calls), leveraged projects/funding versus core funding.



## 1. Community Needs - membership model

Based on a number of assumptions we can offer different services to different users

1. Some organisations have their own institutional repositories, storage and technical expertise (some don't)
2. Some organisations want educational and skills support (not just technical)
3. All want more user engagement and to provide sustained access to content

## Different members, different needs

### 1. Depositors:

- Ingest Collections: store, preserve, display, discover, link. Vast range of data with domain specific needs
- Support leveraged projects: ingest into DRI, use DRI API to re-use content ([Inspiring Ireland](#) and DPC shortlisting)



## Different members, different needs

### 2. Leveraged infrastructures (an instance of DRI):

- Indicated by user needs and institutional requirements
- Re-use of code-base (with licensing)
- Leverages exchequer funding and embeds DRI (knowledge & infrastructure) into institutional workflows
- Pilot study under-way with two universities

## Different members, different needs

### 3. End user:

- Preservation is sustained access to content
- Preservation “out of a box” not “in a box”
- Need to cater to current and future needs
- Tools for user engagement, cross search functionality

## Different members, different needs

### 4. Open data platforms:

- Publish open government data
- No mandate to preserve - need to provide stabilised access
- Publishing is just one step (€3billion for open data pilots?)



## Membership model (until September 2015)

Our membership model is currently under going approval and sign-off.

Builds on legal and policy framework within a federated model based on “trust”

Includes provisions for preservation (within certain storage constraints), functionality (ingest, search, discover, tools to visualisation & analyse).

Priority access to training



## The Challenge

Core funding terminates Sept. 2015

Need for sustained infrastructures & skills

Need to continue building and maintaining DRI

Need to retain staff\*

Project funding

Costed **service** provision

Exchequer core







## We have a compelling argument

DRI is an important infrastructure for our designated community and researchers (citizens, students)

This infrastructure helps other Irish institutions fulfil their open access remit

Partnership and community engagement

Enables Irish researchers to apply for H2020 funding (Irish HEI' will be excluded without this national infrastructure.

[Prof. Rob Kitchin - Business Task Force]



## “Value” of data and software for monetisation

Economic (monetary) versus cultural value

Culture is “priceless” ... for everything else there is  
MasterCard!!

Our ethos is open access to data

We use several open source components to build the  
infrastructure

Number of different business models but we must adhere  
to the values that we have set out



## External factors

Open data, open government - data.gov.ie  
([www.per.gov.ie](http://www.per.gov.ie))

- audit carried out by INSIGHT, many datasets are not available in machine readable formats.
- is preservation included?

H2020 - Ireland needs a national infrastructure

Leveraged projects

## The blended model

Core funding: over 5 years on a sliding scale

Matched funding (philanthropy, Inspiring Ireland)

Membership model

Leveraged research funding

Ingest services

White-label development

In-built archiving costs at source

Next steps are discussions with HEA, stakeholders and other potential funders

[Prof. Rob Kitchin - Business Task Force]

## Conclusion

More to a preservation service than just the infrastructure

Myriad of needs among the community - need to develop services that reflect these

Preservation as a service, that can scale, given community, business, and external factors