Dr Sandra Collins
Director, Digital Repository of Ireland
Royal Irish Academy
Mission

DRI is a trusted digital repository for Humanities and Social Sciences Data

- linking and preserving the rich data held by Irish institutions, with a central internet access point
- Our Cultural & Social Heritage
**DRI Platform**

Access  
Preservation  
Discovery

Federated Archives, Storage
Growing Digital Preservation & Access Policy

Interviews
National Practice Survey
National Steering Committee
National Guidelines
Government adoption

www.oaireland.ie
Metadata

Fig. 5: Metadata standards

- Dublin Core
- EAD/ISAD(G)
- MODS
- MARC 21
- Other
- VRA
- INSPIRE
- METS
- No Standard

% of Institutions (excluding 'don’t know')
Formats

![Bar chart showing formats used by institutions](image)

- RTF
- Word
- PDF
- WAV
- MP3
- QuickTime
- MPEG
- TIFF
- JPEG
- JPEG 2000
- RAW
- XML
- TEI
- Other

% of Institutions (excluding 'don't know')
Global Good Data Practice

Digital Preservation
Data citation, Permanent IDs
Metrics, funding, allowable costs, training
Sustained e-infrastructure
Copyright, IPR, licensing, data protection
Open metadata, open access
Research Data Alliance 2014

Policy, Services, Systems → Practice
Repository

Open source components, custom code engineering

[Logos of OpenNebula, Hydra, Ceph, Apache Solr, Fedora Commons]
OAIS Model
Objects injected into Fedora Commons
Use the Solrizer gem to create the Solr index

Object metadata all CC0
Search will return metadata on all records
Authorization system will restrict access to the objects

Multi-lingual data (English and Irish at the moment)
Indices for each language
User Access

Primarily through the blacklight search interface

Other routes

• Curated collections and virtual galleries
• Georeferenced data – mapping
• Temporal data – timelines
• User defined collections
• DOI references in papers
Linked Logainn is a collaborative project undertaken by the Digital Repository of Ireland (DRI), the Digital Enterprise Research Institute (DERI), Fiontar at Dublin City University, the National Library of Ireland and the Placenames Branch of the Department of Arts, Heritage and the Gaeltacht.

The Linked Logainn project has created a Linked Data version of the authoritative bilingual database of Irish place names logainm.ie, which is being developed by Fiontar in collaboration with the Placenames Branch. The new Linked Logainn dataset provides Irish place name data in a structured, computer-readable format which allows its value to be fully exploited by collaborators, web developers, computer scientists, the heritage community and information professionals. It also encourages the international dissemination of Irish language place names which have been given legal status by the Department of Arts, Heritage and the Gaeltacht.

The Linked Logainn narrative report gives a detailed overview of the project, including an introduction to the collaborating institutions and the Linked Data concept, and outlines the processes undertaken to create the Linked Logainn dataset.
Dublin is the capital and most populous city of Ireland. The English name for the city is derived from the Irish name Dubhlinn, meaning "black pool". Dublin is situated near the midpoint of Ireland's east coast, at the mouth of the River Liffey and the centre of the Dublin Region. Originally founded as a Viking settlement, it evolved into the Kingdom of Dublin and became the island's principal city following the Norman invasion. The city expanded rapidly from the 17th century; it was briefly the second largest city in the British Empire and the fifth largest in Europe. Dublin entered a period of stagnation following the Act of Union of 1800, but it remained the...
Goal of Archive Discovery Project

Linked data based discovery platform
Across multiple RTE Archives, media formats
Enhanced data discovery and delivery
Enhanced workflows, digital practices, tools
Digital Preservation, discovery and access
Two Key Ingredients

1. RDF – Resource Description Framework
   Graph based Data – nodes and arcs
   – Identifies objects (URIs)
   – Interlink information (Relationships)

2. Vocabularies (Ontologies)
   – provide shared understanding of a domain
   – organise knowledge in a machine-comprehensible way
   – give an exploitable meaning to the data
Linked Open Data cloud

Over 200 open data sets with more than 26 billion facts, interlinked by 400 million typed links, doubling every 10 month!
1. Data & System survey
2. Architecture Specification
3. Selection of Pilot Data
4. Retrieval and transformation of data
5. Setup & integration of the metadata repository
6. Metadata enhancement
7. Implementation of content discovery
8. Classification & evaluation of discovery content
9. Demonstrator
10. Performance KPIs
11. Enhanced workflows for content processing
Complex systems, customised software, grown and adapted over time and use.
# Authorisation

<table>
<thead>
<tr>
<th></th>
<th>Public user</th>
<th>Academic researcher</th>
<th>RTÉ researcher</th>
<th>RTÉ journalist</th>
<th>RTÉ Archivist</th>
<th>RTÉ Archives administrator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>View</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>❌</td>
</tr>
<tr>
<td>Amend</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Create</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Delete</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>User mgmt.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>
OUTCOMES FOR RTÉ
OUTCOMES FOR RTÉ

Pilot discovery platform for all media formats

Enabling cross-Institutional, cross-collection curation
Opening up the RTÉ Archives

Through Industry and Research led collaboration

Insight & Background to Project

Bríd Dooley
Head of Archives
RTÉ ARCHIVES

• RTÉ Archives is Ireland’s oldest and largest audiovisual archive
• Expanded and changed as the national broadcaster has grown.
• Employs circa 40 staff supporting over 50,000 research queries p.a
• Supplies up to 10% of original Radio & TV airtime
• Now a major part of RTÉ digital output across all platforms – and growing

Collecting begins 1926
Gramophone Library 1934
Stills Library 1967
Film Library 1968
Sound Archive 1968
TV News Library 1990
Written Archives 1991
TV Sports Library 1995
Film Archive 1996
TV and Radio Archive Projects 2001
RTÉ digital 2012
Our Holdings today – 4 Archival Pillars

1. MOVING IMAGE 210,000 hours

2. AUDIO 150,000 hours

3. PHOTOGRAPHIC 500,000 + selected images

4. DOCUMENTS – 10,000 + new accession -still counting!

Growing at 7% p.@, holdings double every 15 years
Ambition – to open and develop RTÉ Archives as major online digital resource for multiple audiences and user communities.

To date we have produced over 29 curated exhibitions comprising hundreds of moving image & audio segments, documents & images - fully searchable online.
We are already big users of online digital media

Through our website  www.rte.ie/archives

Across daily RTÉ output

On Home page  www.rte.ie  - On this Day

Using Social Media

In special projects ..

As well as a range of partnerships with national cultural heritage institutions and media organisations.
And we know that demand grows as the archive is opened up.
Common challenges to opening up Archives

Large legacy collections on dozens of different playback formats – much of it on discreet carriers.

Four separate library databases /collections with business specific metadata standards for search and query

Broadcast Standard files and formats mixed with Archival requirements in constant state of change

Digital production –archive workflows not yet end-to-end

Interrupted workflows as new systems introduced

Finding those gems! Demands for quick access, search discovery and re-use of our content greater than ever.

Economic constraints of budget & personnel
New tools are needed to semi-automate discovery and curation across our internal databases for better access.

To achieve this we want to work directly with the research community to realise opportunities in linked-data and semantic web technologies inside the broadcast archive specific to our needs.
In Summary – we need to move from Access to Discovery – why?

Because audience & user demands are changing as rapidly as consumer technologies - and we must keep pace

Budgets and staffing shrinking while contend demand and new platforms grow

New tools are needed to aid more automated search and discovery of content both to meet and anticipate expectations & promote the value and use of archives

The search experience we want must not just return accurate results but suggest new pathways of investigation & help to track and document data use through content lifecycle.

RTÉ Archives is now collaborating with two leading research partners in Ireland to investigate how these tools can be developed within RTÉ
Thank you.

Bríd Dooley Head of RTÉ Archives
www.rte.ie/archives
@RTEArchives