Building a national trusted digital repository

IASSIT 2014| JUNE| TORONTO, CANADA

Dr Aileen O’Carroll
Policy Manager
Digital Repository of Ireland
Royal Irish Academy

Dr Sharon Webb
Requirements Analyst
Digital Repository of Ireland
Royal Irish Academy
What is the Digital Repository of Ireland?

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

@anocarroll
Vision

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

- Irish government funded (HEA)
- Royal Irish Academy (lead), National University of Ireland Maynooth, National University of Ireland Galway, Trinity College Dublin, Dublin Institute of Technology, National College of Art and Design
- Partners: academic, cultural, social, government
- Sep 2011 – Sep 2015. Initial grant €5.2M
- Funding currently extended out to 2019
Services

Preservation (TDR, Data Seal of Approval)

Access (use)

Sharing, linking, user tools (reuse)

→ Cultural & Social heritage
Current Status (June 2014)

- Prototype of repository, single and bulk ingest
- Hardware on site; private cloud created
- Initial design of user interface (tools will be bilingual)
- Open Access to Metadata (at a minimum)
- Metadata formats: Dublin Core, MODS, EAD, MARCXML, METS (indexed, searchable, faceted)
- DRI to mint DOIs (PID)
- International Report (follows National one)
- File formats – text and image
- In progress: ingestion guidelines on preparing metadata, audio and video formats, IP/copyright/deposit licenses
Building the Repository through community engagement

National Engagement
International Engagement
Open Source Software
Partnerships & Training

@aaocarroll
National Engagement (cont.)

• Qualitative Interviews with content holders ✓
• Stakeholder Advisory Committee
• National Steering Committee on Open Access
• Demonstrator Projects
• Parallel Projects
Consultation: Interviews with content holders

- Qualitative interviews
- 40 content-holding institutions - current approaches to digital data
- builds engagement, community, trust
- [www.dri.ie/publications](http://www.dri.ie/publications)

@aaocarroll
STAKEHOLDERS

Care for or produce digital content

- Digital repositories*
- University libraries*
- Cultural institutions*
- Social researchers*
- Media organisations*
- Public libraries*
- Government content holders*
- Commercial content holders
- Community and voluntary organisations

Consumers of digital content

- Higher education*
  - Researchers*
  - Lecturers*
- General public
  - Lay researchers
  - Communities of interest
  - Diaspora
- Second-level education
  - Students
  - Teachers
- Policy makers
  - Government
  - NGOs
- Commercial organisations*
Interviewees

An Foras Feasa
Clare County Library
Crawford Gallery
Digital Enterprise Research Institute
Digital Humanities Observatory
Discovery Programme
Dublin City Archives
Dublin City University
Economic and Social Affairs Institute
Health Research Board
Hunt Museum, Limerick
Irish Architectural Archive
Irish Film Institute
Irish Manuscript Commission
Irish Museum of Modern Art
Irish Qualitative Data Archive
Irish Traditional Music Archive
National Archives of Ireland
Nat’l Centre for Tech. in Education
National Folklore Collection
National Irish Visual Arts Library
National Gallery of Ireland
National Library of Ireland
National Museum of Ireland
NUI Galway
NUI Maynooth
Oral History Network of Ireland
Raidió na Gaeltachta
Royal Irish Academy
RTÉ
Trinity College Dublin
University College Cork
University College Dublin
University of Limerick

@aaocarroll
Requirements Engineering – optional slide
Policy Development

Current Practices

Best Practices

Guidelines & National Policy

@aaocarroll
Interview Process

1. Ethical approval
2. Consent for archiving (Test data!)
3. Semi-structured
4. Topic Guide
   • Pre-ingest
   • Ingest
   • Preservation
   • Dissemination
   • Future developments
Formats

Fig. 4: Formats used by institutions

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

% of Institutions (excluding 'don’t know')
Fig. 5: Metadata standards

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

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

@aaocarroll
Thesauri/Vocab/Ontologies

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

- Art & Architecture (Getty)
- Library of Congress
- AACR2
- Dewey Decimal
- Custom
- Other

% of Institutions (excluding ‘don’t know’ or N/A)
Digital Asset Management Systems

38% of institutions interviewed had own repositories
Database / Where do you keep records?

**Fig. 7: Database formats and systems**

- CSV
- Excel
- Access
- MySQL
- Oracle
- SQL Server
- FileMaker Pro
- PostgreSQL
- Other

<table>
<thead>
<tr>
<th>Database Format</th>
<th>% of Institutions (excluding ‘don’t know’ or N/A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSV</td>
<td>11</td>
</tr>
<tr>
<td>Excel</td>
<td>33</td>
</tr>
<tr>
<td>Access</td>
<td>33</td>
</tr>
<tr>
<td>MySQL</td>
<td>42</td>
</tr>
<tr>
<td>Oracle</td>
<td>11</td>
</tr>
<tr>
<td>SQL Server</td>
<td>22</td>
</tr>
<tr>
<td>FileMaker Pro</td>
<td>11</td>
</tr>
<tr>
<td>PostgreSQL</td>
<td>11</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
</tr>
</tbody>
</table>

@aaocarroll
Challenges to Policy – Open Access

- Copyright (Humanities and cultural heritage)
  - Shifting Legal Frameworks
  - Copyright restrictions
  - Orphan works

- Data protection (Social sciences)
  - Anonymity
  - Embargos
  - Access restrictions
  - Data destruction
DRI Policy Responses - Access

- National Principles for Open Access Policy Statement
  - [http://tiny.cc/IRLOpenAccess](http://tiny.cc/IRLOpenAccess)

- Open metadata

- Identified metadata standards for ingest into DRI

- Interoperability with Europeana
Conclusion

- Old issues
- New issues, new TDRs?

User Engagement

- Training
- Skills sharing
- National Policy development

@aaocarroll
Training

Workshops/Seminars:

• Linked Data, Data Modelling, Visualisation (and all about DH)
• ‘Translation’ workshops by request – e.g. archivists, librarians
• Digital Preservation in a Day
• Open Access
• Public Events (e.g. Researchers’ Night Curate-a-thon)

Partnerships with:

• Digital Preservation Coalition (trust, digital pres. for beginners)
• CENDARI (European collaborative research infrastructure)
• Central Statistics Office - curation roadshow (4th quarter 2013)

INSPIRING IRELAND http://www.inspiring-ireland.ie

2014: Documentation on ingestion guidelines, hands-on workshops

@aaocarroll
Contact us!
@dri_ireland
@aaocarroll
Aileen.ocarroll@nuim.ie

Preview DRI: http://vimeo.com/95853217

Inspiring Ireland: http://vimeo.com/89125000