



Digital Repository of Ireland  
*Taisclann Dhigiteach na hÉireann*

## DRI: Preservation Planning Case Study Getting Started in Digital Preservation

Digital Preservation Coalition | November 2013 | Dublin, Ireland

Dr Aileen O'Carroll  
Policy 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

## Development of a Preservation Plan

1. Policy Frameworks
2. Stakeholder Engagement
3. Requirements engineering
4. Strategy Development and Testing

## Services

Preservation

Access (use)

Sharing, linking, user tools (reuse)

→ *Cultural & Social heritage*

## 1. Digital Preservation –Robust Policy Framework



Policy Framework

Data Seal of Approval  
(DSA)

<http://datasealofapproval.org/en/>

ISO 16363  
(TRAC checklist)

## Data Seal of Approval - Producers

1. The data producer deposits the research data in a data repository with **sufficient information** for others to assess the scientific and scholarly quality of the research data and compliance with disciplinary and ethical norms.
2. The data producer provides the research data in **formats** recommended by the data repository
3. The data producer provides the research data together with the **metadata** requested by the data repository

## Data Seal of Approval - Repositories

4. The data repository has an explicit mission in the area of digital archiving and promulgates it
5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.
6. The data repository applies documented processes and procedures for managing data storage
7. The data repository has a **plan for long-term preservation** of its digital assets

## Data Seal of Approval - Repositories

8. Archiving takes place according to **explicit workflows** across the data life cycle
9. The data repository assumes responsibility from the data producers for access and availability of the digital objects
10. The data repository enables the users to utilize the research data and refer to them
11. The data repository ensures the **integrity** of the digital objects and the metadata
12. The data repository ensures the **authenticity** of the digital objects and the metadata
13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS

## Data Seal of Approval - Consumers

14. The data consumer complies with access regulations set by the data repository
15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in higher education and research for the exchange and proper use of knowledge and information
16. The data consumer respects the applicable licenses of the data repository regarding the use of the research data

## DRI & Data Seal of Approval I

Guide to Data Preparation and Archiving  
Depositing Data Guide

Format Policy (including policy on non-preferred Formats)  
Metadata Policy

Ingest Policy or Data Deposit Guides  
Ingest Procedures/ Quality Control & Quality Assurance

Procedures that include

- Policy on changing and updating archived data
- Policy on Version Control

## DRI & Data Seal of Approval II

Depositor Agreement or Licenses  
Policy on Copyright, Ownership, IP and Data Protection  
Policy on Sensitive Data & Ethical Statement (including  
Takedown)

Access Policy that includes

- \* Breaches Policy
- \* DOI Policy

Access/End User Licenses that includes agreement on  
use, re-use, access and ownership

## DRI & Data Seal of Approval III

**Preservation Policy** that includes

- Data Integrity Strategy
- Back up strategy
- Migration Strategy
- Audit Strategy
- Disaster Recovery Plan

Procedural Document on Data Lifecycle / Repository  
Operations Policy

Policy on Trusted Digital Repository (including statement  
on OAIS)

# More Detail? ISO 16363TRAC

**3.3.5 The repository shall define, collect, track, and appropriately provide its information **integrity measurements**.**

## **Supporting Text**

This is necessary in order to provide documentation that it has developed or adapted appropriate measures for ensuring the integrity of its holding.

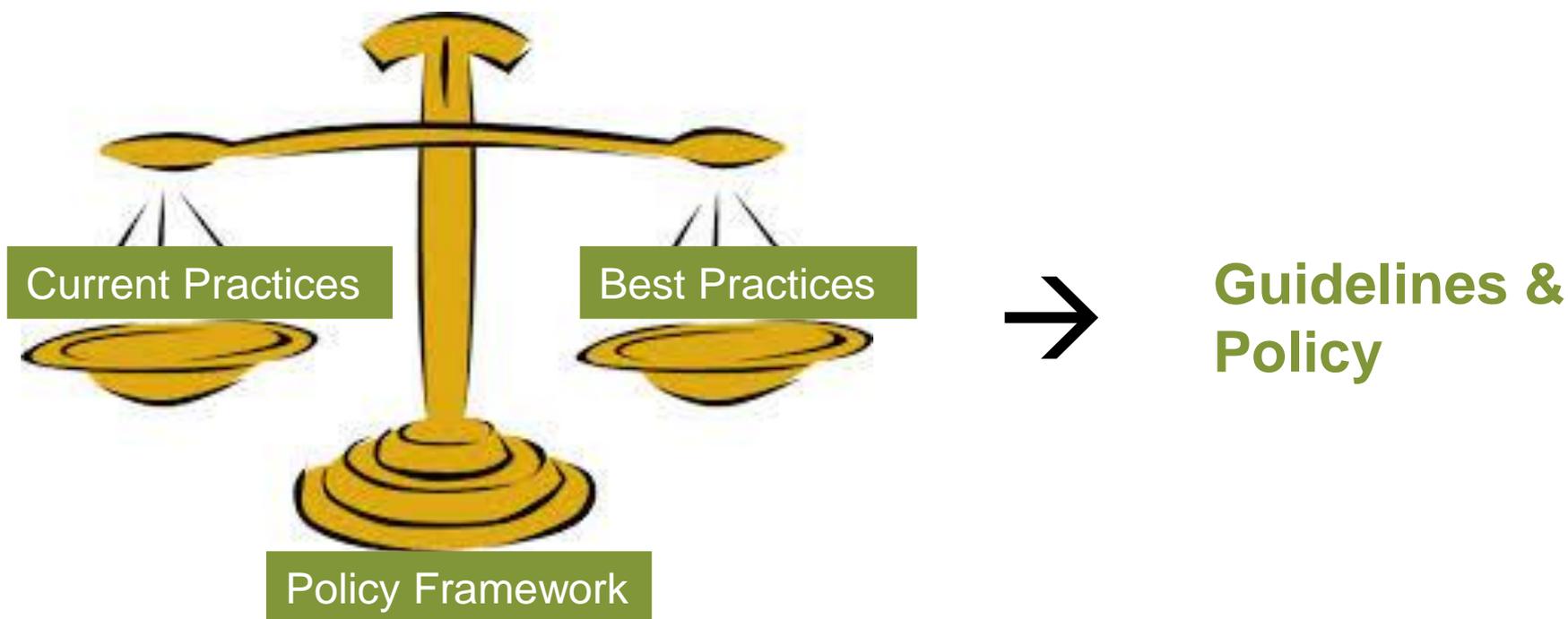
## **Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement**

Written **definition or specification** of the repository's integrity measures (for example, computed checksum or hash value); **documentation of the procedures** and mechanisms for monitoring integrity measurements and for responding to results of integrity measurements that indicate digital content is at risk; an **audit process** for collecting, tracking and presenting integrity measurements; **Preservation Policy** and **workflow documentation**.

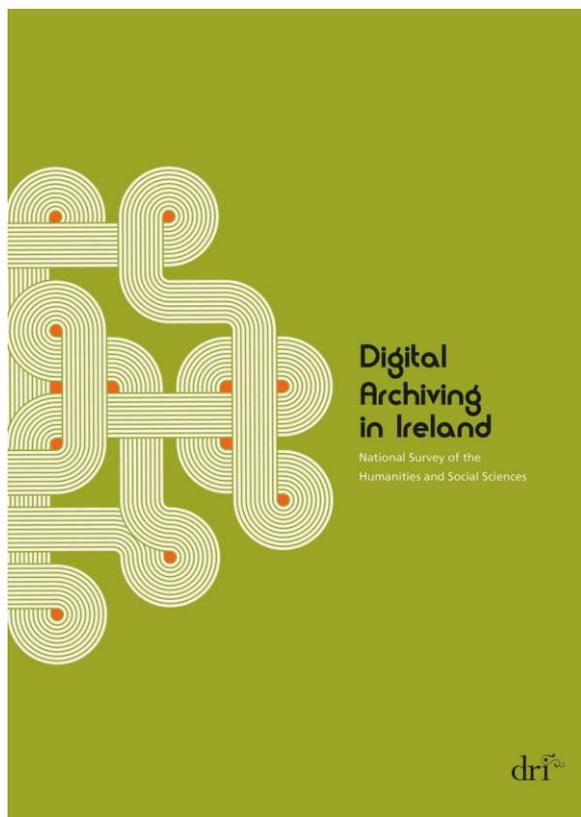
## **Discussion**

The mechanisms to measure integrity will evolve as technology evolves. The repository may provide documentation that it has developed or adapted appropriate measures for ensuring the integrity of its holdings. If protocols, rules and mechanisms are embedded in the repository software, there should be some way to demonstrate the implementation of integrity

## 2. Digital Preservation through community engagement



## Consultation: Interviews with content holders



- Qualitative interviews
- 40 content-holding institutions - current approaches to digital data
- builds engagement, community, trust
- Launch of National Report in Oct 2012: *Digital Archiving in Ireland: National Survey of the Humanities and Social Sciences*
- [www.dri.ie/publications](http://www.dri.ie/publications)

# 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

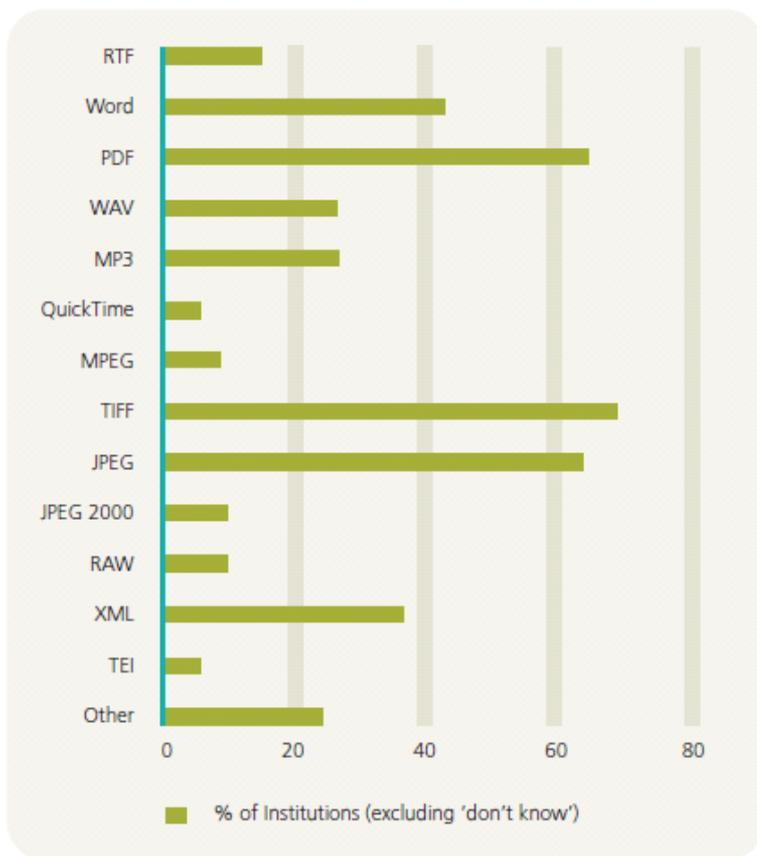
Irish Traditional Music Archive  
National Archives of Ireland  
Nat'l Centre for Tech. in Education  
National Folklore Collection  
National Irish Visual Arts Library  
National Gallery of Ireland  
National Library of Ireland  
National Museum of Ireland  
NUI Galway  
NUI Maynooth  
Oral History Network of Ireland  
Raidió na Gaeltachta  
Royal Irish Academy  
RTÉ  
Trinity College Dublin  
University College Cork  
University College Dublin  
University of Limerick

## 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



### Textual:

Preferred: PDFa, rtf, txt, xml

Accepted: PDF, DOC, DOCx

### Audio:

Preferred: WAV, BWAV

Accepted: MP3

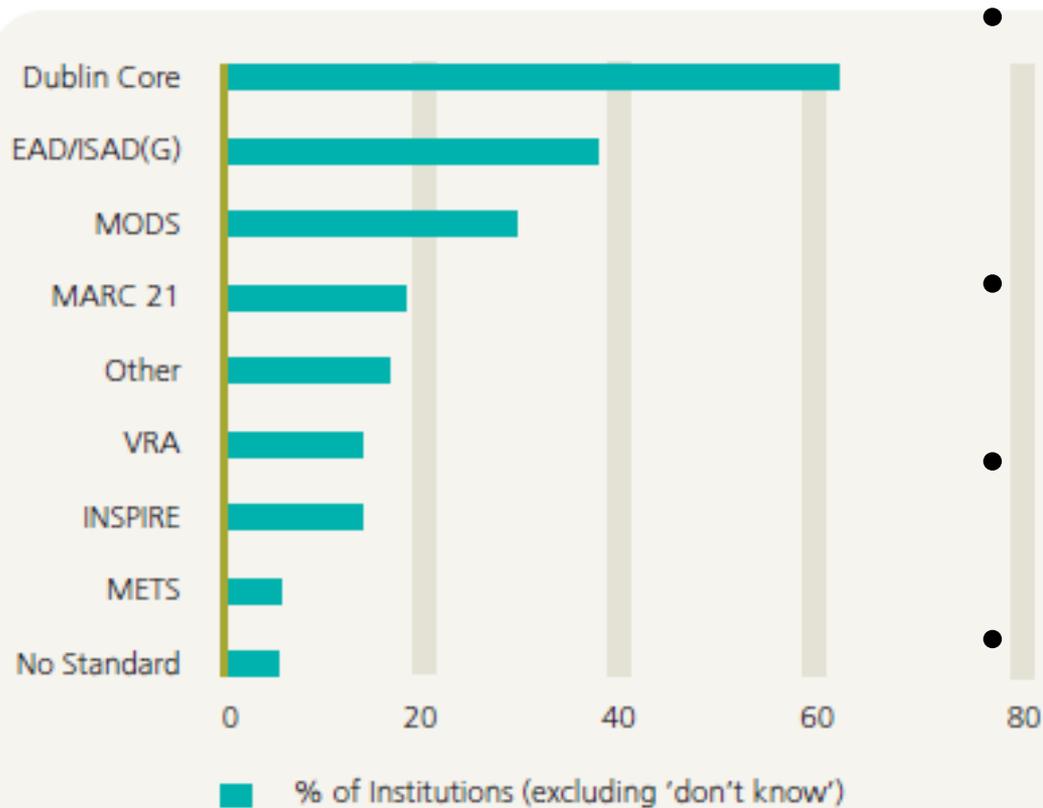
### Image:

Preferred: TIFF,

Accepted: JPEG.

## Metadata

Fig. 5: Metadata standards



- Dublin Core (simple and qualified)
- MARC
- MODS
- EAD

## Digital Asset Management Systems

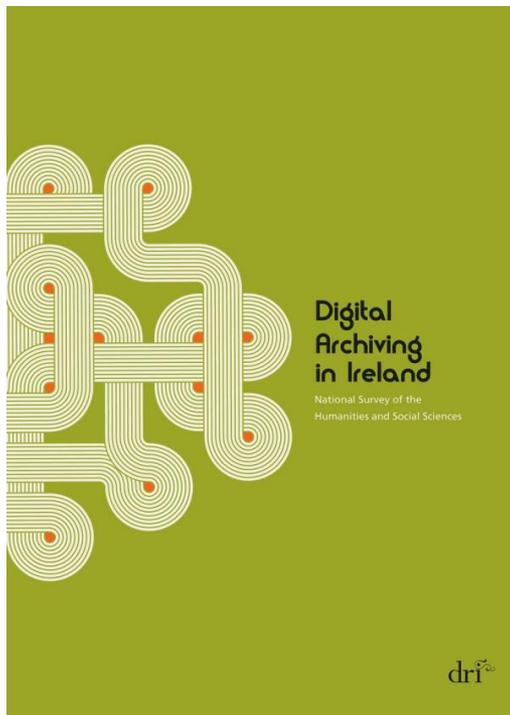
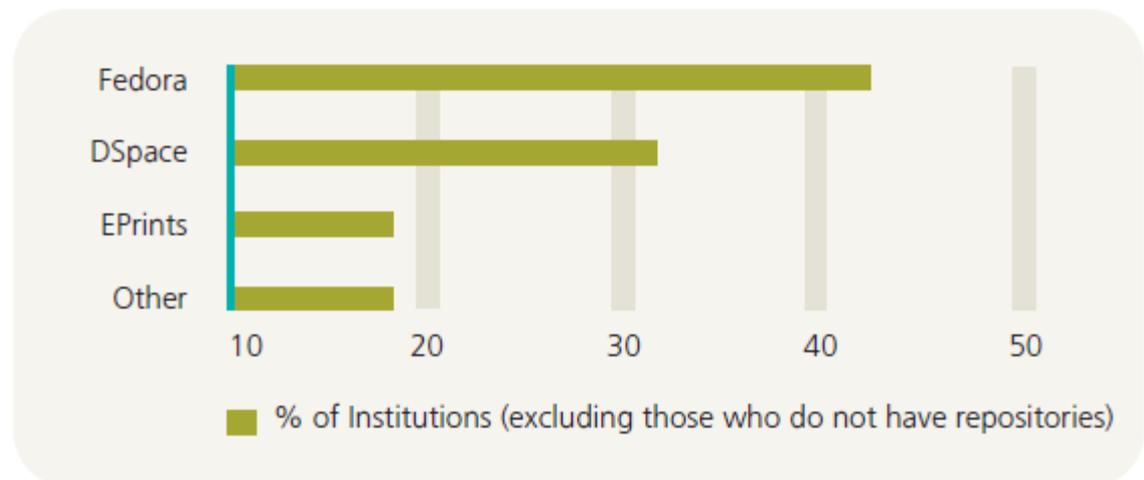


Fig. 8: Digital asset management systems

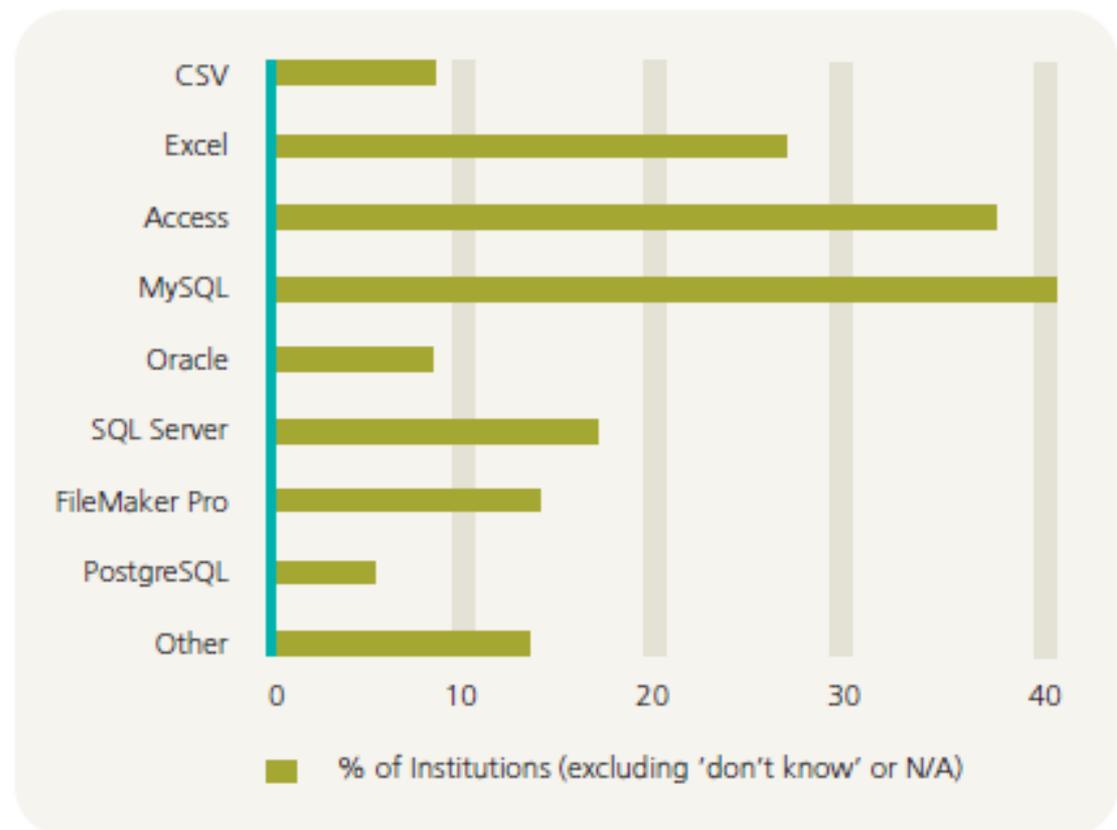


38% of institutions interviewed had own repositories

## Database / Where do you keep records?



Fig. 7: Database formats and systems



## 4. Requirements Engineering

RE specifies what the system or product should do. It ensures that the system is built upon authentic user requirements.

Defines a projects objectives and helps to produce resources/projects/software that considers both the context (the user, the environment of use, the problem domain) and the software development effort.

Concerned with real-world goals, functions and constraints.

## 4. REQ- 34 Data Preservation

The system must "identify and manage the risks to its preservation operations and goals associated with the system infrastructure".

1.1 It must "employ technology watches or other technology monitoring notification systems"

1.2 It must "have procedure in place to monitor and received notifications when hardware changes are needed"

1.3 It must "have procedures in place to monitor and receive notifications when software changes are needed"

1.4 It must allow a user, in accordance with their access rights, migrate from one format to another in response to changes or obsolescence in software or media (see REQ-21).

## 4. REQ-60 Data Integrity

The system must check the integrity of all communications and data held in the repository

- 1.1 It must check the integrity of digital objects at delivery and access point
  - 1.1.1 It must calculate a checksum and compare with checksum created at point of ingestion.
- 1.2 It must check the integrity of all data held in the repository periodically.
- 1.3 It must prevent unauthorized corruption of all data held in the repository.

## 4. REQ-62 Audit Trail

### REQ-62 Audit trail

The system must manage and save all information related to internal user and external user and system interaction with the repository in accordance with current legislation.

- 1.1 It must track all changes made to a digital object.
- 1.2 It must track all changes made to a collection.
- 1.3 It must track all changes made to user access permissions.
- 1.4 It may track all user log in and log outs.
- 1.5 It must track all virus scans and integrity checks that are carried out by the system, including intermittent system checks as well as at point of ingestion.

### Related REQ-17.4 Checksum

The system must create and record a checksum for all digital objects at point of ingestion.

## 4. Strategy Development and Testing

1. Identifying Issues and documentation required (eg below)
2. Identifying what is currently supported Fedora/Hydra
3. Identifying absences (TRAC checklist)
4. Testing with demonstration projects

### ISSUES

#### **Audit**

- User (access permissions)
- Object (interactions with objects)
- System
- Organisation/Business

#### **Preservation**

- Bit-level, object, system, organisation
- data integrity
- Migration
- Disaster Recovery

## Developing a Preservation Plan

1. Policy Frameworks (DSA, ISO 16363)
2. Stakeholder Engagement (know your designated community)
3. Requirements engineering
4. Strategy Development and Testing

# Contact us!

@dri\_ireland

@aaocarroll

[Aileen.ocarroll@nuim.ie](mailto:Aileen.ocarroll@nuim.ie)