

**REBIUN**  
**VIII Workshop sobre Proyectos Digitales:**  
**“La preservación digital: la memoria futura”**

***La preservación digital  
y las universidades:  
estado de la cuestión.***

**Alice Keefer, [alice.keefer@ub.edu](mailto:alice.keefer@ub.edu)**  
**Profesora colaboradora, Univ. de Barcelona**

**Murcia, 21 octubre 2008**

# Introducción

La evolución de la preservación digital en los 5 últimos años.

- Revistas científicas digitales: Portico, CLOCKSS, E-Depot
- Pautas, estándares: TRAC, PREMIS
- Herramientas: JHOVE, GDFR, PRONOM, DROID
- Proyectos europeos: PrestoSpace, Caspar, Planets, Shaman

Todo ello gracias a la colaboración interinstitucional e/o internacional

# Introducción

Identificación automática de los elementos técnicos:

*DROID: Digital record object identification*

<http://droid.sourceforge.net/wiki/index.php/Introduction>

*JHOVE: JSTOR/Harvard object validation environment.*

<http://hul.harvard.edu/jhove/>

*Global digital format registry.*

<https://collaborate.oclc.org/wiki/gdfr/about.html>

*Pronom: The technical registry.*

<http://www.nationalarchives.gov.uk/pronom/#>

# Introducción

La evolución de la preservación digital en los 5 últimos años: investigación y práctica:

cuestiones técnicas → estrategias,  
migración/emulación;

cuestiones institucionales → fiabilidad,  
viabilidad/sostenibilidad, certificación;

cuestiones prácticas → metadatos, derechos de autor, análisis de riesgos.

# Introducción

La **viabilidad institucional** (*institutional viability*): la transición requerirá la capacidad técnica, económica y material para asumir las nuevas competencias.

La **sostenibilidad económica** (*economic sustainability*): La capacidad de sostener el nivel de apoyo necesario –sobre todo el económico– en el futuro.

\* \* \* \* \*

*"La falta de preparación ('readiness') de la organización, y no la tecnología, es el principal inhibidor de programas de preservación digital."*

**Kenney, 2004**

# Introducción

La evolución de la preservación digital en los 5 últimos años: acontecimientos relacionados:

- Movimiento Open Access;
- Creación de los repositorios institucionales

# Introducción

## ¿Qué es la preservación digital?

- acciones destinadas a mantener la accesibilidad de los objetos digitales a largo plazo (**Directrices para la preservación**)
- ...actividades necesarias para asegurar el acceso continuado a materiales digitales hasta cuando sea necesario –no obstante los obstáculos que representan los fallos en los soportes o los cambios tecnológicos. (**DPE Handbook**)
- ...actividades gestionadas necesarias para asegurar el mantenimiento a largo plazo de la cadena de bits y la accesibilidad continuada del contenido. (**Trusted digital repositories**)
- ...acciones necesarias para asegurar un acceso continuado y pertinente a la información digital durante el tiempo que sea requerido y para cualquier finalidad legítima. . (**Waller/Sharpe**)

# La preservación del contenido de los RIs

¿Qué material hay en los RIs?

- resultados de la investigación (tesis, e-prints);
- objetos de docencia y aprendizaje;
- datos no elaborados;
- fondos digitalizados;
- material administrativo.

¿Qué material tienen que preservar a largo plazo?



# La preservación del contenido de los RIs

## La teoría:

*"an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution."* **Lynch, 2003**

*"...most would include long-term preservation of digital materials as a key function of IRs."* **Pathfinder report, 2004**

*"...there is a growing interest among research funding agencies in data management, curation and archiving that is not necessarily closely coupled to the open access debates....IRs will, in our view, be an important vehicle for addressing these data curation obligations."*

**Lynch/Lippincott, 2005**

*"Preservation and trustworthiness are undoubtedly the foundation of any repository..."* **Chavez et al, 2007**

# La preservación del contenido de los RIs

## La realidad:

El RI de Cranfield University: “*is more concerned with access than preservation....*” **Bevan, 2007**

“*no repositories surveyed had a formal preservation policy to guide decisions on the questions raised in the survey...*” **Hitchcock et al, 2007a**

# La preservación del contenido de los RIs

En la UK, hay poco consenso sobre el papel de los RIs en cuanto a la preservación:

- ¿cada institución será responsable de su propia producción digital?
- ¿se crearán servicios especializados nacionales que se encargarán de estas operaciones?

Posiblemente **para el 2010** se habrá definido los papeles y las responsabilidades para preservar a largo plazo los recursos depositados en los RIs.

# La preservación del contenido de los RIs

Criterios tradicionales para tomar la decisión sobre la preservación a largo plazo:

Principalmente los factores de

- valor
- pertinencia
- uso
  
- Otros condicionantes: misión, disponibilidad de recursos humanos, económicos, materiales, obligaciones legales o contractuales.

# La preservación del contenido de los RIs

## Criterios nuevos para los recursos digitales:

- que la institución tenga pleno derecho a manipular los datos para asegurar su acceso en entornos informáticos del futuro;
- que el recurso sea de un formato legible actualmente y previsiblemente en el futuro;
- que si el recurso esté en un soporte gestionable para su transferencia y/o almacenamiento;
- que el recurso disponga de documentación, incluyendo los metadatos.

# La preservación del contenido de los RIs

## Selección de recursos para su preservación

¿Qué formatos? ¿qué versiones? ¿qué material adicional incluir?

¿Qué atributos se quieren preservar?

- datos, funcionalidad
- apariencia, esencia (*Look and feel*)

La decisión dependerá de la misión institucional, las necesidades de la comunidad de usuarios, la capacidad técnica/ tecnológica institucional y los recursos disponibles.

# La preservación del contenido de los RIs

## Texto

- Microsoft Word
- PDF

## Otros de Microsoft Office

- Microsoft PowerPoint
- Microsoft Excel
- Microsoft Access

## Video/Animation files

- Flash
- QuickTime

*Table 5: File formats used by  
Academics at Loughborough  
University*

## Ficheros basado en web

- HTML

## Imágenes

- GIF
- JPEG
- BMP

## Otros

- CAD
- FileMaker Database
- Hot Potatoes
- LaTeX

[Bates, Melanie et al. \*Digital lifecycles and file types: final report.\* JISC; Loughborough University, 2006.](#)

# Los derechos de autor

El RI necesita obtener permiso de los titulares para realizar determinadas acciones:

- Difusión
- Reproducción
- Modificación / transformación

Además este permiso posiblemente tendrá que extenderse a terceros – por ejemplo, proveedores de servicios de preservación.



# Los derechos de autor

M.I.T. -- Licencia de depósito:

*"You agree that MIT may, without changing the content, translate the submission to any medium or format for the purpose of preservation. You also agree that MIT may keep more than one copy of this submission for purposes of security, back-up and preservation."*

University of California Libraries: requerimientos:

*"the submitter must have the right to authorize the deposit for preservation and copying."*

# Los metadatos en los RIs

Son necesarios para permitir:

- el acceso a los documentos depositados
- su gestión y preservación a largo plazo

descriptivos: datos sobre el contenido intelectual, que ayudan a identificar y localizar un recurso;

administrativos: datos necesarios para la gestión de un recurso: mantenimiento, almacenamiento, entrega; incluyen los aspectos técnicos y los relacionados con los derechos.

estructurales: datos que se refieren a la estructura del recurso y a los elementos que lo integran.

# Los metadatos en los RIs

Creación de metadatos:

Similar a la catalogación / indización:

- Trabajo profesional laborioso,

Pero a diferencia de estas operaciones:

- Elementos difíciles de conocer
- Gran volumen de documentos digitales generados
- Alto porcentaje de documentos únicos
- Necesidad de actualizar los metadatos al largo de la vida del documento

# Los metadatos en los RIs

LIFE<sup>2</sup>

## LIFE Model v1.1: Stages and Elements

Lifecycle Stage	Creation or Purchase	Acquisition	Ingest	Metadata Creation	Bit-stream Preservation	Content Preservation	Access
	...		Selection	Quality Assurance	Re-use Existing Metadata	Repository Admin	Preservation Watch
...		Submission Agreement	Deposit	Metadata Creation	Storage Provision	Preservation Planning	Access Control
...		IPR & Licensing	Holdings Update	Metadata Extraction	Refreshment	Preservation Action	User Support
Lifecycle Elements	...	Ordering & Invoicing	Reference Linking		Backup	Re-ingest	
		Obtaining			Inspection		
		Check-in					

7

Los elementos que afectan el coste de la preservación digital.

# Los metadatos en los RIs

## ¿Cómo crear los metadatos?

- Personal del RI (o de la biblioteca): catalogadores, etc.;
- Creadores / depositores;
- Generación automática:
  - Del texto
  - De fuentes externas (bbdd, Reference Manager, etc.)

# Los metadatos en los RIs

## Univ. of Hull – RepoMMan Project

Hay que automatizar la creación de metadatos: *it is not viable for the metadata to be entirely human-generated*

- Metadatos “contextuales” – extraídos de los sistemas administrativos institucionales
- Metadatos descriptivos – extraídos del documento mediante el programa “iVia” (Univ. California - Riverside)
- Metadatos técnicos – generados por JHOVE

**Green et al, 2007**

# Los metadatos en los RIs

## Proyectos JISC sobre metadatos:

- AIR (Automated Archiving for an Institutional Repository): Univ. of Wolverhampton, September 2007 - March 2009
- Digital Plait: Univ. of Aberystwyth, May – December 2008
- EIDeR: Kings College London, April 2008 – March 2009

# Los metadatos en los RIs

EIDeR: Enhanced ingest to digital e-research repositories (Kings College London)

[www.jisc.ac.uk/whatwedo/programmes/programme\\_rep\\_pres/interoperabilitydemos/eider.aspx](http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/interoperabilitydemos/eider.aspx)

Desarrollar interfaces “máquina-a-máquina” para la extracción automática de metadatos, la preservación, los derechos de acceso, depósito y captura.



# La fiabilidad de los repositorios

*Trustworthy Repositories Audit and Certification.*

(TRAC) NARA, OCLC, CRL, Feb. 2007.

Criterios para identificar los repositorios digitales capaces de almacenar, migrar y dar acceso a fondos digitales, centrados en 3 aspectos:

- infraestructura institucional
- gestión del objeto digital
- tecnologías, infraestructura técnica y seguridad

# La fiabilidad de los repositorios

Otras iniciativas de evaluación de depósitos  
fiabiles:

*Catalogue of criteria for trusted digital repositories.*  
NESTOR, 2006.

<http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf>

*DRAMBORA: Digital Repository Audit Method Based on  
Risk Assessment*

<http://www.dcc.ac.uk/tools/drambora/>

# La fiabilidad de los repositorios

Externalización de servicios:

Varios modelos:

- gestión técnica
- gestión administrativa:
- gestión del acceso

Ex. Digital Preservation Repository de la California Digital Library (CDL) <http://www.cdlib.org/inside/projects/preservation/dpr/>

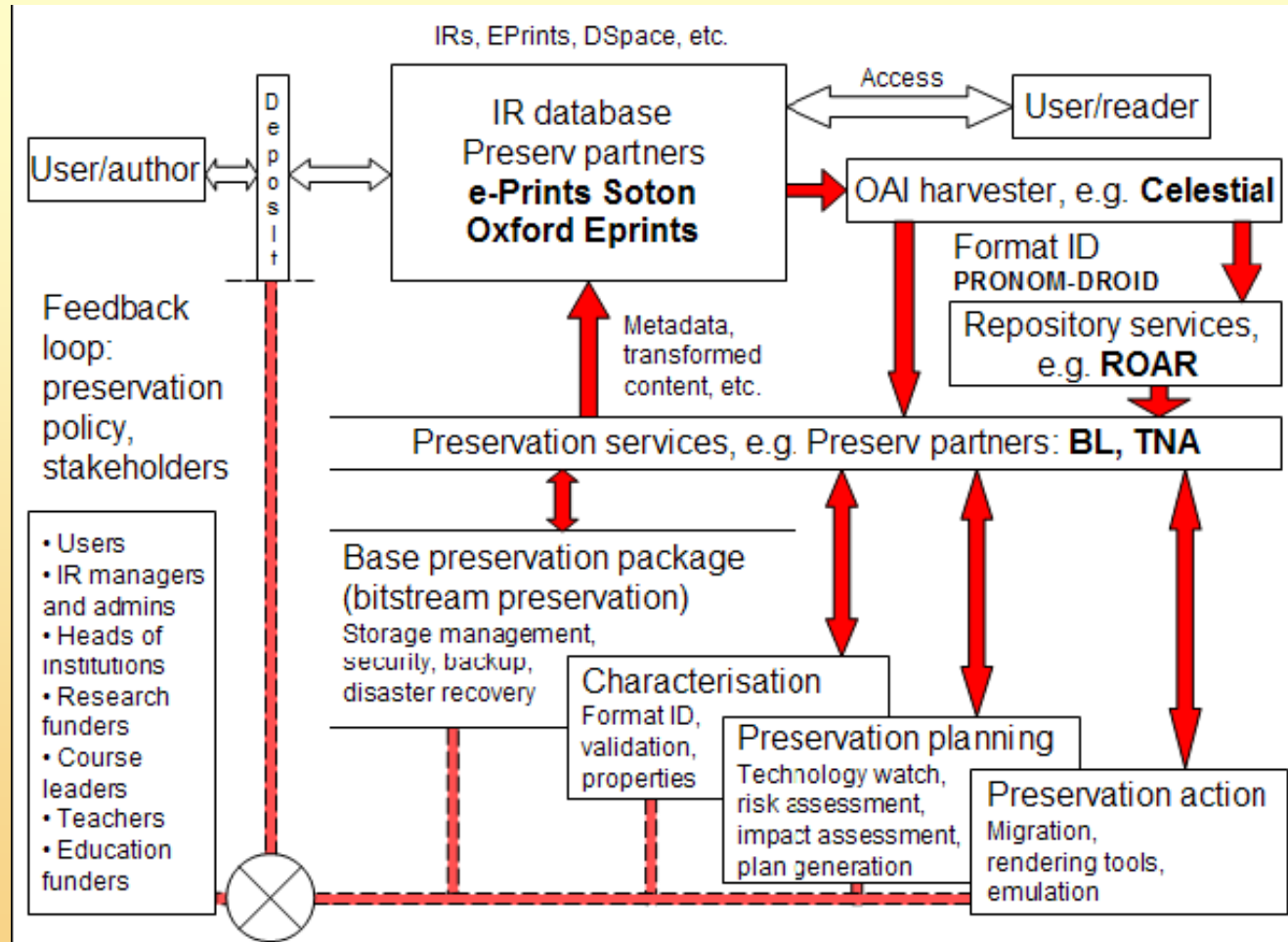
# La fiabilidad de los repositorios

PRESERV project (2005-07); Univ. Southampton:  
<http://preserv.eprints.org/>

- Preservación a largo plazo en los RIs:  
identificación de servicios externos de preservación;
- Incorporación de interfaces con los servicios en el software utilizado en los RIs (comenzando con EPrints)

Hitchcock et al, 2007a

# La fiabilidad de los repositorios



# La fiabilidad de los repositorios

PRESERV project:

*PRONOM-ROAR*

<http://trac.eprints.org/projects/iar/wiki/Profile>

Un enlace desde el directorio de RIs ROAR al software PRONOM (The National Archives) para permitir la identificación y verificación de los formatos en los RIs.

Ej:

[http://roar.eprints.org/?action=search&query=southampton  
&sa=Search](http://roar.eprints.org/?action=search&query=southampton&sa=Search)

[http://roar.eprints.org/?action=profile&url=http%3A%2F%2F  
eprints.soton.ac.uk%2F](http://roar.eprints.org/?action=profile&url=http%3A%2F%2F<br/>eprints.soton.ac.uk%2F)

# La fiabilidad de los repositorios

PRESERV project :

Cambios al software EPrints:

incorporación de un módulo para guardar información sobre la historia del objeto digital: cambios ocurridos, acciones realizadas, etc.;

- implantación de aplicaciones para empaquetar y difundir los datos para su entrega a servicios externos (utilizando METS o DIDL);
- modificación del proceso de depósito para permitir la captación de metadatos de preservación;
- incorporación de una opción para que los autores indiquen los derechos asignados a los gestores de los servicios.