

Repository preservation services: divisible, viable and sustainable?

Steve Hitchcock

Preserv 2 Project

Intelligence Agents Multimedia Group,
School of Electronics and Computer Science (ECS),
Southampton University

Digital repositories: Dealing with the digital deluge,
JISC conference, Manchester, 5-6 June 2007

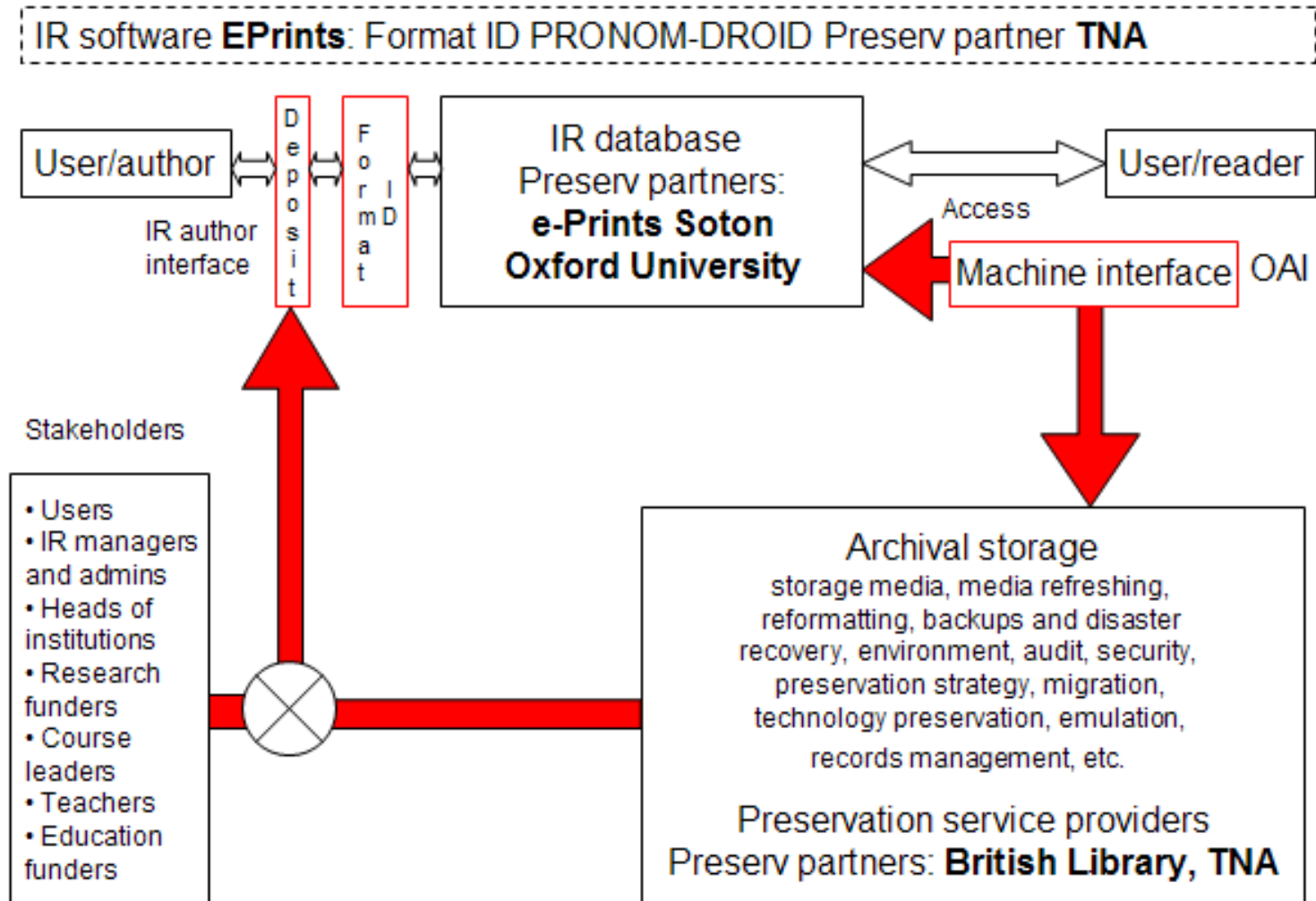


JISC

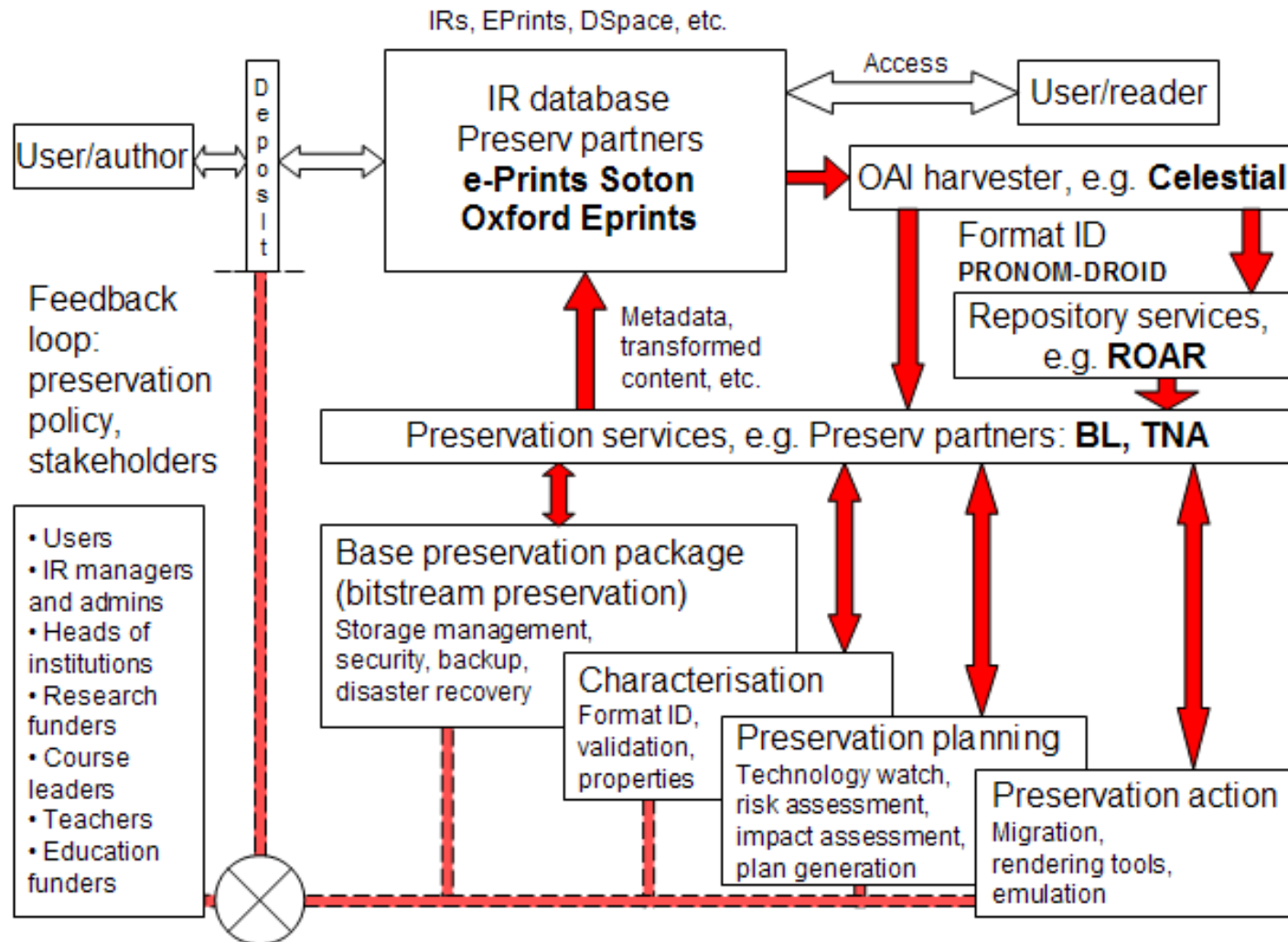
A
the national archives

LIBRARY
HSLIRB

Preserv preservation service schematic: original



Preserv preservation service schematic: latest



Preserv: a migration

Preserv, from February 2005 to January 2007
within the **JISC 4/04 programme** Supporting Digital
Preservation and Asset Management in Institutions

Preserv 2, just approved, likely timeframe June 2007 (tbc)
to December 2008 (18 months)
funded in response to **JISC circular 4/06, September
2006: Repositories**

Preserv partners: changing places

Preserv Preservation service provider:

- The British Library

Software provider: The National Archives, Southampton University

Data providers: Southampton University, Oxford University

Preserv 2 Preservation service providers:

- The British Library
- The National Archives
- Oxford University

Other providers as before

Preserv: selected outputs

Video

- *Preserv Us! The story of the Preserv project by the people behind it* <http://preserv.eprints.org/video-coverpage.html>

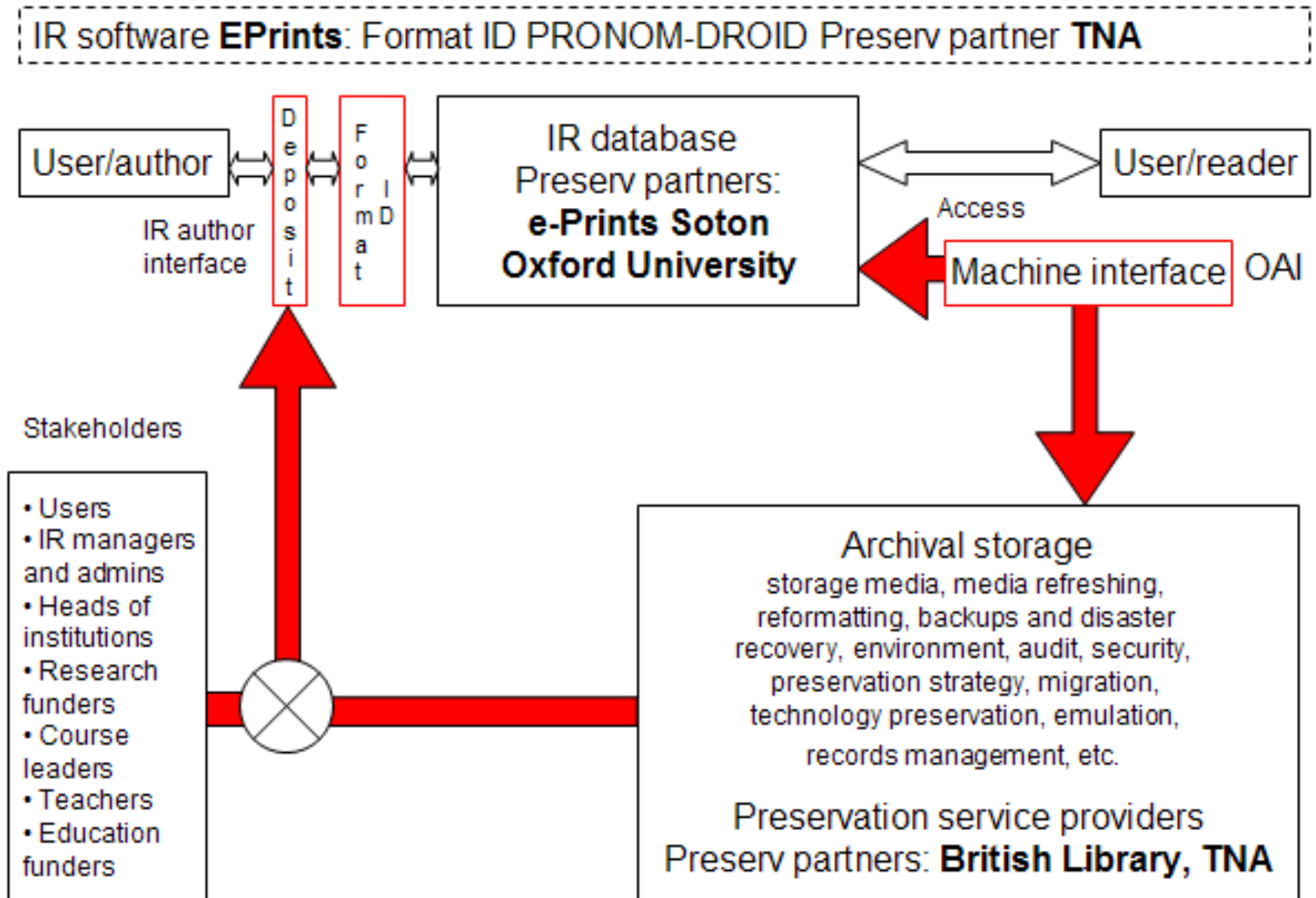
Papers

- Digital Preservation Service Provider Models for Institutional Repositories: Towards Distributed Services, *D-Lib Magazine*, Vol. 13, No. 5/6, May/June 2007
- Survey of repository preservation policy and activity <http://preserv.eprints.org/papers/survey/survey-results.html>

Web page

- Preserv Format Profiling: PRONOM-ROAR: An illustrated guide <http://trac.eprints.org/projects/iar/wiki/Profile>

Preserv schematic: original



Some reassurance for IRs

- Preservation is a scary, long-term business.
- IRs are relatively new (since 2000), characterised by growth and diversity, principally to provide immediate open access to deposited content.
- There is nothing to be scared about at the moment. The problem is manageable while IRs are new.
- It won't be scary if IRs develop appropriate policies and engage with preservation service providers at an early stage. Preservation isn't long-term, it's progressive, with simple and practical steps taken now paying dividends later.

Preservation: responsibilities of repositories

- Repository preservation must start with policy
- Repository policy is concerned with all aspects of repository management: content strategy (mandates!), collection policy, rights, etc., see OpenDOAR Policies Tool <http://opendoar.org/tools/en/policies.php>

- Preservation strategy will emerge from this analysis

BUT

- Repositories can only know the requirements and scale of the preservation task with a fully formed policy
- To fulfil institutional requirements, the policy needs institutional backing (a true *institutional* repository)

How are repositories doing so far?

- Does the repository have any existing policy on preservation?

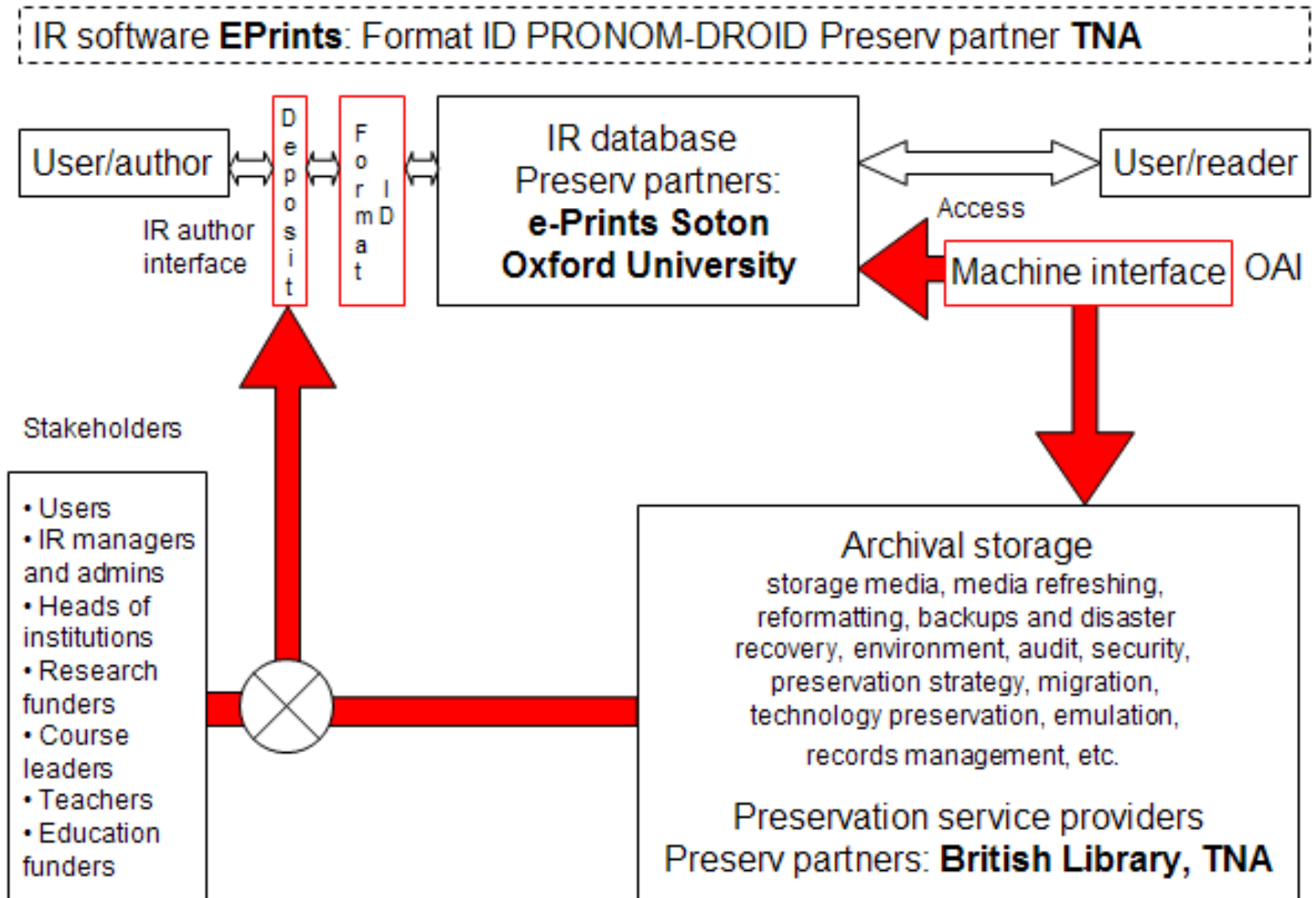
Yes 1 No 20

- Does the repository have a policy on submission file formats?

Yes 11 No 4 (no reply 4)

- There will be a session on preservation planning at the **RSP Repositories Summer School 2007**, Dartington College, Totnes, Devon, June 27 - 29. Sign up now!

Preserv schematic: original



Format profiling using PRONOM and ROAR

1. A search is performed in ROAR for archives containing *soton* (Southampton domain)

2. Clicking the *Formats* button generates a format summary for all matched archives

3. Clicking a bar shows a breakdown of all files identified as that format (e.g. *PostScript 2.0*) and associated OAI records

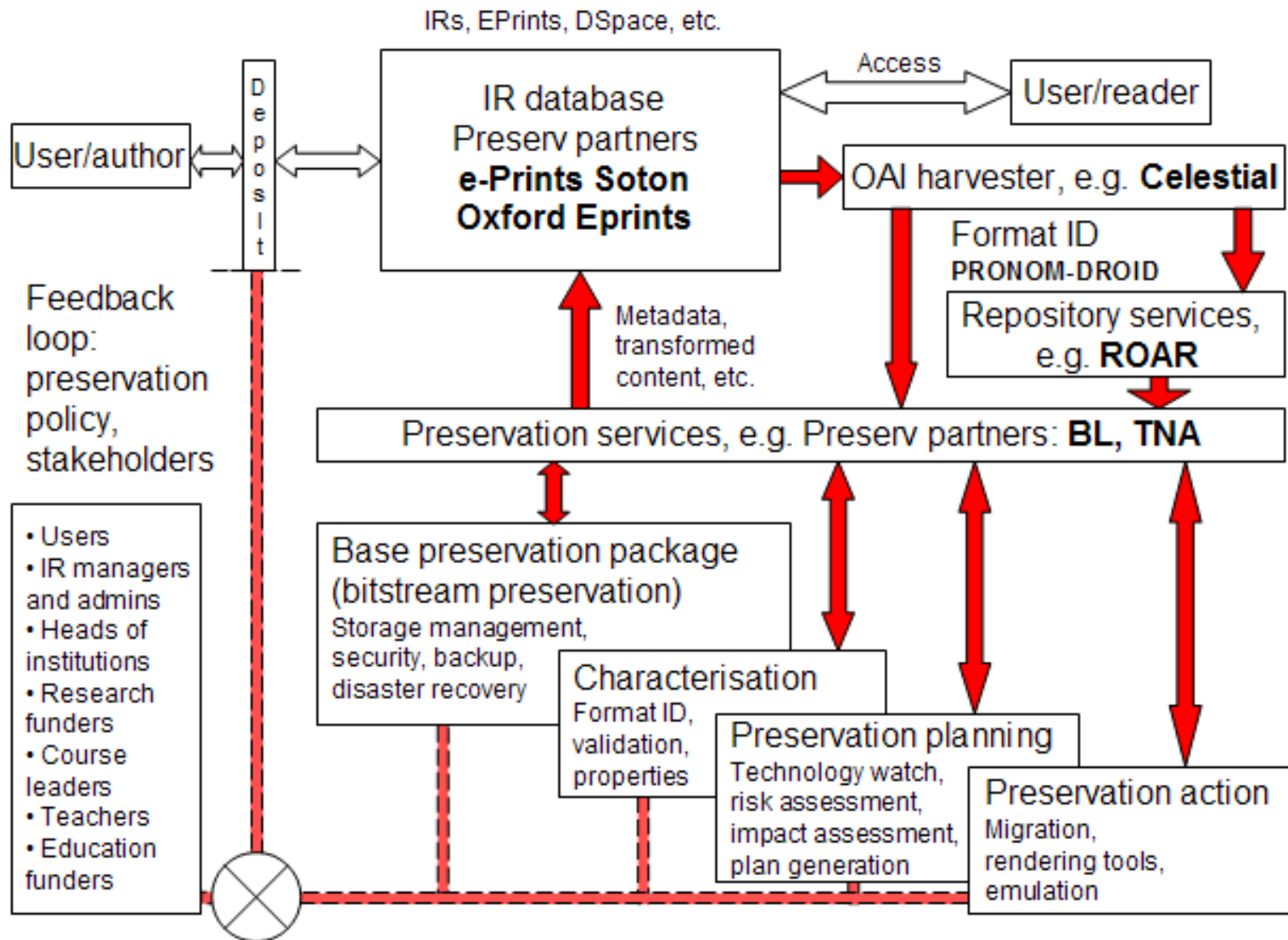
4. Clicking the OAI record identifier shows the Dublin Core record (from which the files were located)

Note: While some PostScript documents also have a PDF version, many do not: this could be the basis for an *alerting service* to prompt the migration of preservation-unfriendly formats

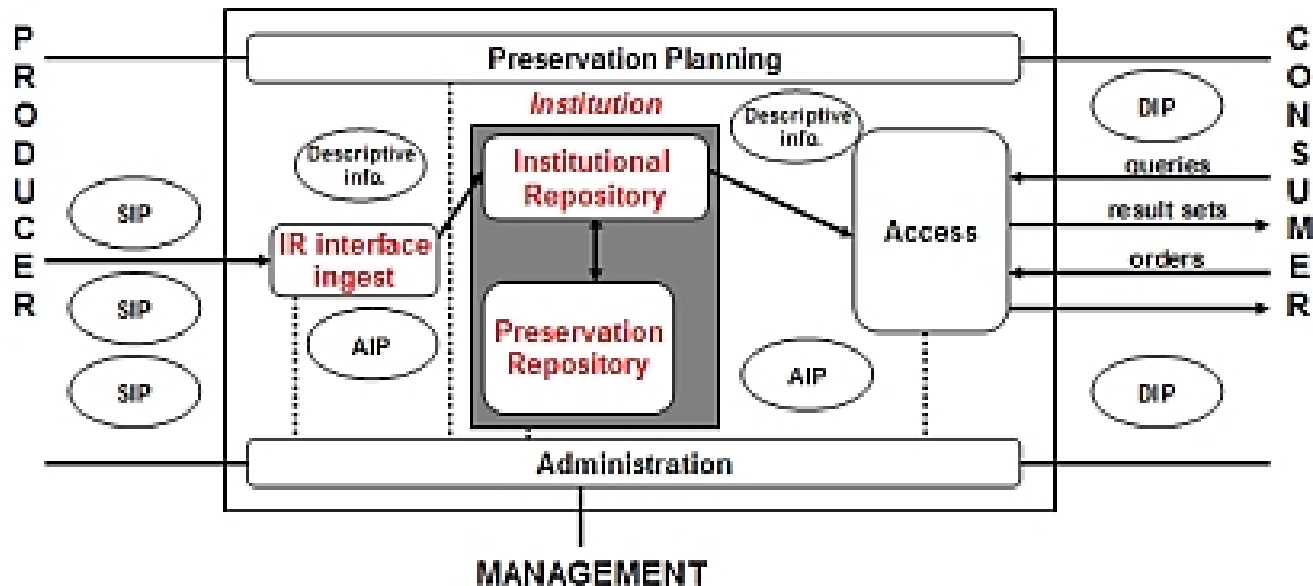
Format	Count
Portable Document Format (1.5)	202
Portable Document Format - Archival (1)	1
Microsoft Word for Windows Document (97-2002)	1
Graphics Interchange Format (1989a)	1
Microsoft Powerpoint Presentation (97-2002)	1
Portable Document Format (1.1)	1
OLE2 Compound Document Format	1
GZIP Format	1
Hypertext Markup Language (4.01)	1
Portable Document Format (1.6)	1
JPEG File Interchange Format (1.01)	1
Hypertext Markup Language (4.0)	1
ZIP Format	23
Rich Text Format (1.0)	19
PostScript (3.0)	17
Adobe Illustrator	14
Binary Interchange File Format (BIFF) Workbook (8)	9

Date	1997-09-01
Resource Type	Conference or Workshop Item
Resource Identifier	http://eprints.ecs.soton.ac.uk/1910/
Format	pdf http://eprints.ecs.soton.ac.uk/1910/03/weiser97b.pdf
Format	ps http://eprints.ecs.soton.ac.uk/1910/04/weiser97b.ps

Preserv schematic: latest



Institutions might have a preservation repository



Institutional model, Figure 2c from Digital Preservation Service Provider Models for Institutional Repositories, *D-Lib Magazine*, May/June 2007 <http://www.dlib.org/dlib/may07/hitchcock/05hitchcock.html>

Preserv 2: what we will do

- **Preservation planning:** TNA will develop preservation planning, including technology watch and risk assessment.
- **Exemplar preservation action services** (e.g. migration tools) that respond to the outcomes of preservation planning.
- **Migration of objects and metadata** between repositories (e.g. EPrints to Fedora): Oxford University will develop a generic harvester to gather objects and metadata through various interfaces and perform the necessary transformations.
- **Develop ROAR and associated OAI services** to support distributed preservation (Southampton University).
- **Evaluating the market potential.**

Moving to a new concept: distributed preservation services?

- We need to look beyond the idea of a 'black box' preservation service
- Services *might* be based on lightweight, interacting distributed Web services
- Who will provide these services?
- What coordination is required between services? Is that where client-facing service providers will emerge?
- What services can the market sustain?

See DPC Featured project interview: Preserv, 25 July 2006

<http://www.dpconline.org/graphics/join/preserv.html>