

PRESERV Preservation Eprint Services

PRESERV: Preservation Services for OAI-Compliant Repositories

<http://preserv.eprints.org/> Jessie Hey, Tim Brody, Steve Hitchcock and Leslie Carr
University of Southampton, UK

The OAI-PMH has become the de-facto standard for exposing metadata. In the PRESERV project (1) we have explored new models for enabling the digital preservation of and long-term access to content in Institutional Repositories (IRs). We envision digital preservation being achieved through simple preservation services working with standards-based, interoperable repository software. As support for the OAI-PMH matures so repositories are providing more robust mechanisms to access their content through OAI, e.g. community standards for using Dublin Core, support for METS or DIDL. Based on this, we have developed an exemplar File Format Profiling tool in the Registry of Open Access Repositories (ROAR), utilizing OAI and PRONOM DROID (2). PRONOM-ROAR is a first step to preserving digital content through simplifying content file format management for IR Managers by providing file format profiles and alerts.

1. PRESERV (Preservation Eprint SERVICES) is a JISC-funded project led by the University of Southampton with partners at the British Library, Oxford University and The National Archives.

2. PRONOM-DROID (Digital Record Object Identification) is a software tool to perform the automated batch identification of file formats. DROID is the first in a planned series of tools developed by The National Archives under the umbrella of its PRONOM technical registry service. PRONOM is a database of file format metadata and signatures for identifying files based on their bit streams. The collaboration with PRESERV helped provide valuable feedback to improve the DROID tool in the first instance and an open source version was released in August 2006.

The DROID tool was used to provide a trial PRESERV service to repository managers through the Registry of Open Access Repositories (ROAR). ROAR already provided simple graphs to track the growth of metadata records in individual repositories using OAI-PMH. The PRESERV service was added to ROAR by downloading all files and then identifying them using the DROID tool. A 'PRESERV Profile' interface provides a break-down of file formats by repository. An email alerting service sends periodic emails on request to repository managers informing them of the number of new records and the format of any associated digital objects in their repositories. This could form the basis for a *technology watch* service that would alert managers to files at risk of obsolescence.

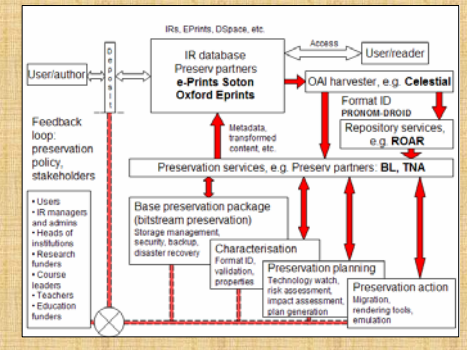
With institutional repository content increasing in breadth and depth we expect more unusual formats to be deposited, particularly, in the Humanities. While there is further work to be done to refine the output of email alerts, and to explore additional preservation services, open and collaborative services such as this show promise for simplifying the management of repositories in the longer term. Other services such as JHOVE and PANIC are exploring complementary solutions and the Library of Congress provides expertise on additional formats.

Preservation Services in the Repository Lifecycle

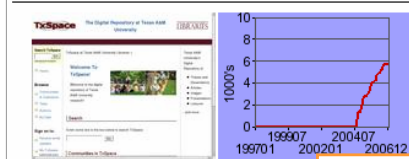
Scenario: Digital lifecycle begins with author creation and deposit of paper or data content into the institutional repository (IR). Growing number of IRs with expanding content.

Problem: Authors and IR editorial staff typically have content management skills, but preservation expertise is more thinly spread.

Solution: Many third-party preservation services. Adapt IR software to disseminate content to centres of preservation excellence, to provide preservation features to IRs.



The Digital Repository at Texas A&M University (5686 records)



1. A search performed in ROAR for archives matching 'texas'

2. Clicking the *Preserv Profile* link generates a format summary for the given repository
Registered on 2006-05-04
(OAI) Total OAI Records: 5686 [Preserv Profile]
100% freely accessible fulltext (estimate)

Format	Total Files
[No files found]	0
Portable Document Format (1.5)	865
Tagged Image File Format (3)	619
Portable Document Format (1.4)	607
Portable Document Format (1.3)	323
Portable Document Format - Archival (1)	160

3. Clicking a bar shows a breakdown of all files identified as that format (e.g. PDF 1.5) and associated OAI records

Format	Count	Sample Files	Date
Portable Document Format (1.5)	865	oai:txspace.tamu.edu:1969.1/137...m/1969.1/137/1/etd-tamu-2003C-CECN-Reach1-1.pdf	2006-05-11 02:46:26
Tagged Image File Format (3)	619	...m/1969.1/137/2/etd-tamu-2003C-CECN-Reach1-2.zip	2006-05-11 02:46:26
Portable Document Format (1.4)	607	oai:txspace.tamu.edu:1969.1/137...m/1969.1/137/1/etd-tamu-2003C-CECN-Reach1-1.pdf	2006-05-11 02:46:26
Portable Document Format (1.3)	323	...m/1969.1/137/2/etd-tamu-2003C-CECN-Reach1-2.zip	2006-05-11 02:46:26
Portable Document Format - Archival (1)	160	oai:txspace.tamu.edu:1969.1/138...969.1/138/1/etd-tamu-2004A-MEEN-Rodriguez-1.pdf	2006-05-11 02:46:26
Binary Interchange File Format (BIFF)	1	...969.1/138/3/etd-tamu-2004A-MEEN-Rodriguez-2.xls	2006-05-11 02:46:26
Audio/Video Interchange	1	Binary Interchange (ab01r)\desktop\etd-tamu-2003C-CECN-Reach1-2.zip	2006-05-11 02:46:26
Quicktime	7	DecisionTree... FlowDiagram... 2.xls	
Windows Media Video	4	DecisionTre... eDocument... Masterchec... Physical and e-Document...	
Waveform Audio	4		
Portable Document Format (1.1)	3		
Binary Interchange File Format (BIFF) Workbook	5		
RealAudio Video	1		
MPEG-1 Video Format	1		
Microsoft Word for Windows Document	1		

4. Some formats may require administrative investigation, for example what a Zip file contains: *in this case a group of data sets associated with a research paper*



<http://roar.eprints.org/>
e.g. ROAR for Switzerland in Google Earth



A Simple Preservation Service: File Format Analysis and Alerting Service