

Smart Storage and Preservation: How Digital Repositories Can Participate

Sally Rumsey: University of Oxford
Steve Hitchcock: University of Southampton
Neil Jefferies: University of Oxford
Adrian Brown: The National Archives

DLF Forum: November 2008



Jeffrey Keefer <http://www.flickr.com/photos/jeffreykeefer/773540725/>
CC licence: Attribution-Non-Commercial-Share Alike 2.0 Generic





Crystal ball by Hamachi! CC license. **Attribution-Non-Commercial-No Derivative Works 2.0 Generic**
Available at <http://www.flickr.com/photos/mawari/2091456761/>

[All Project RSS](#) 
[RSS Admin](#)

[Project News](#) Latest...



From open to smart storage: Preserv 2 goes international

An important progression from open to smart storage will be presented in two papers to be given by Preserv at international meetings this month. Open storage combines open source software ... [more](#)

Project Partners



School of Electronics
and Computer Science



Project Advisors



Preserv: Inspiration and Provision of Preservation Services for Digital Repositories

[Project Overview](#)

[Objectives & Outcomes](#)

[Outputs and Services](#) 

[Papers & Presentations](#) 

[Project News](#) 

[Preserv Blog](#)

[People](#)

"It is important to build the concept of preservation from the outset. In the digital era, the 'outset' for most new research and educational materials will be the institutional repository."*

* JISC Circular 40/06, note 10

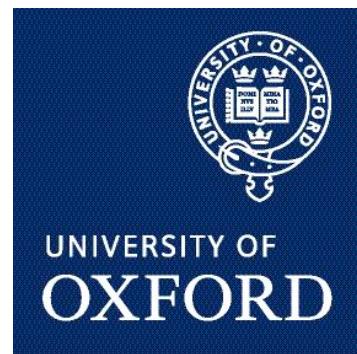
About the project

Fast facts

- **When:** Preserv 2 from July 2007 to December 2008 (Preserv 1 ran from February 2005 to January 2007)
- **Who:** Southampton University (School of Electronics and Computer Science), The National Archives, The British Library and Oxford University (Library Services, Systems and Electronic Resources Service) ... [collaborating people](#).
- **Prior related work by the partners:**
 - Southampton: [EPrints repository software](#), [TARDIS project](#), [a sustainable multidisciplinary institutional repository](#)
 - TNA: [digital preservation at TNA](#), [PRONOM](#), [the file format registry](#)
 - BL: [Digital Object Management \(DOM\) Programme](#)
- **Funded by** JISC, within the programme [Supporting Digital Preservation and Asset Management in Institutions](#)
- **Keywords:** digital libraries, institutional repositories, preservation, preservation services, EPrints, PRONOM

Coverage of Preserv

May 1, 2007 Project video. [Preserv Us! The story of the Preserv project by the people](#)



JISC

[home](#)[about us](#)**what we do**[funding opportunities](#)[publications](#)[news](#)[events](#)[contact us](#)[Programmes](#)[Capital programme](#)[Projects funded](#)[Calls for projects roadmap](#)[email this page](#)

Capital programme



JISC theme(s): [e-Research](#), [e-Learning](#), [Network](#), [e-Resources](#), [Information environment](#), [e-Administration](#), [Access management](#)

Committees: [JISC Learning and Teaching committee](#), [JISC Network committee](#), [JISC Support of Research committee](#), [JISC Content Services committee](#), [JISC Integrated Information Environment committee](#), [JISC Organisational Support committee](#)

The capital programme is additional funding of £90 million over three years to enhance the network infrastructure (SuperJANET 5), to digitise key resources for the academic community, and to support the development of e-learning; e-infrastructure; virtual research environments; users and innovation; and repositories and preservation. Funding for SuperJANET 5 has been provided by all JISC's funding partners. However, the remaining activities have been funded by the Higher Education Funding Councils for England and Wales ([HEFCE](#) and [HEFCW](#)).

Funding opportunities 2006/2007

See projects funded under the capital programme

See the [road map](#) for calls, [summary information](#) for the background to the programme and [progress of the capital programme](#).

July 07: Enterprise Architecture, e-Research: e-Infrastructure and Users and Innovation

[Circular 02/07](#) released 27 July 2007 (Deadline: 2 October 2007).

April 07: e-Learning, Repositories and Preservation and e-Research: e-Infrastructure

[Circular 01/07](#) (Deadline: 21 June 2007) [Circular briefing day](#) (9 May 2007, previously known as a Town Meeting) held on for those interested in submitting proposals - FAQs and copies of the presentations are available to download. [See supporting information](#)

September 06: e-Learning, Users & Innovation, Repositories & Preservation and e-Research: e-Infrastructure and Virtual Research Environments



3rd Annual WePreserve Conference
Nice, France
28-30 October 2008

Welcome to Planets

Planets, Preservation and Long-term Access through Networked Services, is a four-year project co-funded by the European Union under the Sixth Framework Programme to address core digital preservation challenges. The primary goal for Planets is to build practical services and tools to help ensure long-term access to our digital cultural and scientific assets. Planets started on 1st June 2006. This website makes available project documentations and deliverables as Planets progresses so that these can be shared with the libraries, archives and digital preservation community.

The strong Planets consortium brings together expertise across Europe from national libraries and archives, leading research universities and technology companies. Coordinated by the British Library, the partners are:

- The British Library
- The National Library of the Netherlands
- Austrian National Library
- The Royal Library of Denmark
- State and University Library, Denmark
- The National Archives of the Netherlands
- The National Archives of England, Wales and the United Kingdom
- Swiss Federal Archives
- University of Cologne
- University of Freiburg
- HATII at the University of Glasgow
- Vienna University of Technology
- Austrian Research Centers GmbH
- IBM Netherlands
- Microsoft Research Limited
- Tessella Support Services Plc

» Home

- » About Planets
- » News
- » Events
- » Publications
- » Software
- » Contact Us

News

Starting out: Preserving Digital Objects - Principles and Practice

18th June 2008

DPE, Planets, CASPAR and nestor are delighted to announce that they will be delivering the latest in a series of collaborative training initiatives to introduce the Principles of Digital Objects Preservation. This event will be held in Prague, Czech Republic from October 13-17 2008. [\[More\]](#)

Digital Preservation Planning: Principles, Examples and the Future with Planets

16th June 2008

PLANETS and the Digital Preservation Coalition (DPC) will be holding a joint briefing day: Digital Preservation Planning: Principles, Examples and the Future with Planets, on 29th July 2008, at the British Library Conference Centre, St Pancras, London. [\[More\]](#)



Preservation & Storage Formats for Repositories

support@rsp.ac.uk

Overview

Formats matter if digital content is to be accessible now and preservable in the longer term. Institutional Repositories (IRs), which provide access to and store digital objects produced by many creators, will need to manage a range of formats. This briefing paper explains how formats affect preservation, considers which formats repositories should use for deposit and storage, and describes the practical steps repositories can take to produce an initial preservation plan.

How are formats used?

Digital documents are produced, in one form or another, using an application program such as a word processor. These documents are encoded with information to represent characters, layout and other features. The rules of the encoding are defined by the chosen format of the document. Applications are often closely tied to formats, e.g. Microsoft Word can be used to produce the document (.doc) format, Adobe Acrobat produces the portable document format (.pdf). These may not be the only formats that an application can produce, e.g. Word can also produce Rich Text Format (.rtf), and formats may not be exclusive to one application.

Why are formats important for preservation?

Problems with application-specific formats can arise when users try to open a digital document without access to the application that was used to create it, or without the correct version of an application. This is most likely to happen when opening a document created by someone else. This problem increases over time, that is, it becomes harder to open documents in their original format if the application has changed or no longer exists.

If applications and formats can change over time, it follows that some risk becoming obsolete. This is why formats are a primary focus for preservation actions, and why repositories need to be aware of the formats of the digital objects they store.

Which deposit formats should an IR allow?

There are many different types of digital objects (e.g. texts, images, videos), and many different applications for producing them. There are also different views on which formats are the most 'preservable', and therefore which formats a repository should allow to be deposited. There is one format that an IR should always commit to obtaining: the author's source format. That is, the version produced by the author directly from the application used at the time of completion.

The most common example of deviating from this approach is a requirement for authors to deposit PDF, which is not an authored format - it is created by converting from another format. By requiring authors to submit the source format for preservation, the repository can then convert to its preferred presentation format, which could be PDF, if that is different from the source format. It is likely this conversion can be automated and, in the process, documented.

Which formats should repositories commit to support in the long-term?

The key phrase that describes the ideal longer-term storage format is *open standard*, meaning the specification is freely available and implementable. Consequently it is more likely that applications to view and use such formats will be available at any given time, since viewers can be developed by the wider community of users with an interest in the format, and not just the original application developer. Open standard formats include OpenDocument format (ODF), an XML file format for electronic office documents.



ora

OXFORD+UNIVERSITY
RESEARCH+ARCHIVE

Search Detailed Search

Browse
Contribute
Technical
Help

Oxford University Research Archive (ORA) contains research publications and other research output produced by members of the University of Oxford. Content includes copies of journal articles, conference papers, theses and other types of research publications. The full text of many of these items is freely available to be used in accordance with copyright and end-user permissions.

Oxford University Research Archive is a growing repository of Oxford research publications and is therefore not a complete record of the research output from the university. Members of the University of Oxford may deposit items in Oxford University Research Archive. For more details contact us.

[Disclaimer and Data Protection statement](#) [Accessibility statement](#) [Download Copyright guide \(100KB PDF\)](#) [Take-down procedure and complaints \(30KB PDF\)](#) [Contact](#)

Site Copyright © 2008 Oxford University Research Archive All rights reserved.



F MO

forced migration online a world of information on human displacement

Home Contact us About us Policy Site Map Search

Regional Resources Thematic Resources Digital Library Journals Web Catalogue Organizations Podcasts Videos Photos Working papers Events FAQ Blog Discussion List Feedback Donate

REFUGEE STUDIES CENTRE

home > digital library > search >

Digital Library - Search

The FMO digital library is a unique online resource. It contains over 9,000 full-text documents and journal articles in electronic format which can be searched, read online and printed as required. The documents include both recent and historical grey (unpublished) literature and research materials. The digital library contains full-text articles from back issues of key journals in the field. For more background information please refer to our [Case Study](#) and [Scope Policy](#) documents.

To search **both** documents and journals, enter your search terms below. To limit your search to a specific collection of documents or journals, use the "Advanced Search" option. For more information, please refer to the [Search Tips](#).

Submit a document Advanced Search Search Tips ?

Search

Full text Sort by Relevance Go



University of Oxford Faculty of History

Search website Find

Home

Most of the experience of humanity is contained in the past. Medieval and modern history has been studied at Oxford for longer than at almost any other university: a Regius Professor of Modern History was first appointed in 1724, and undergraduate examinations began in 1850.

University of Oxford Mathematical Institute

Mathematical Institute EPrint Server Home About Browse Search User Area Help

The Mathematical Institute Eprints Archive

Welcome to The Mathematical Institute Eprints Archive

This is the eprint server for the [Mathematical Institute](#), [University of Oxford](#).

Search records:

Search



TNA's Seamless flow approach to preservation

1

**Format
identification
and
characterisation**

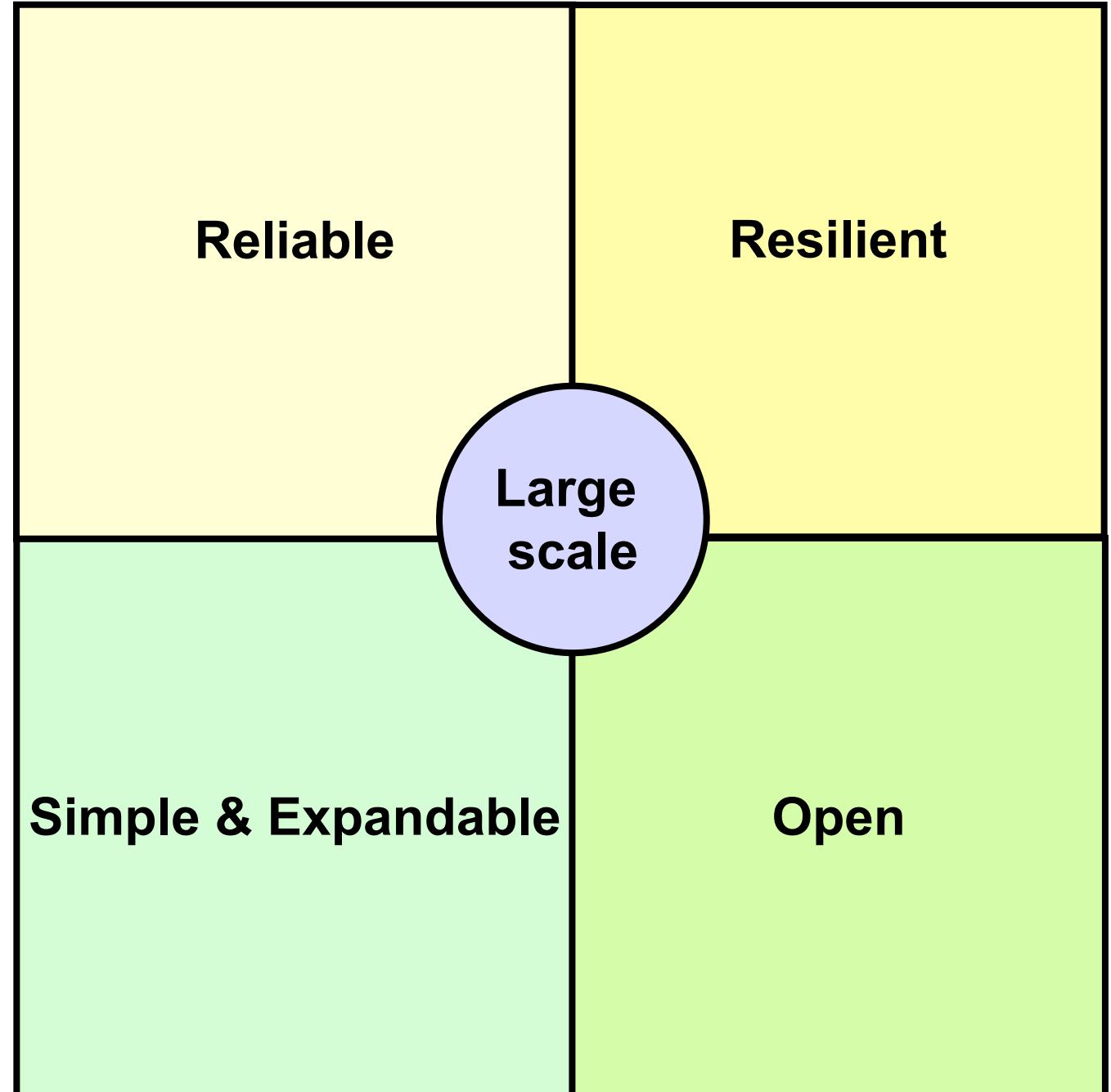
2

**Preservation
planning
and
technology
watch**

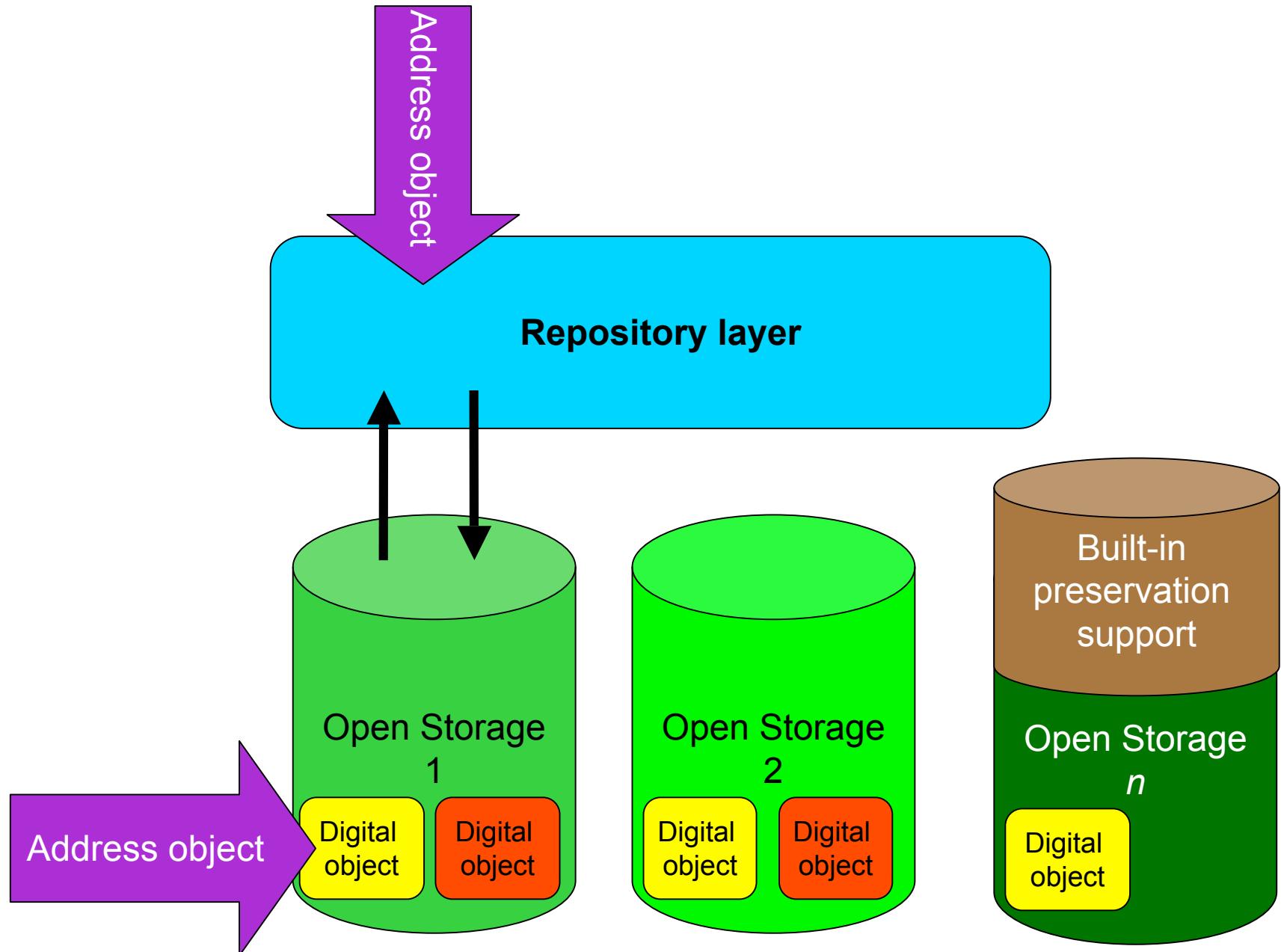
3

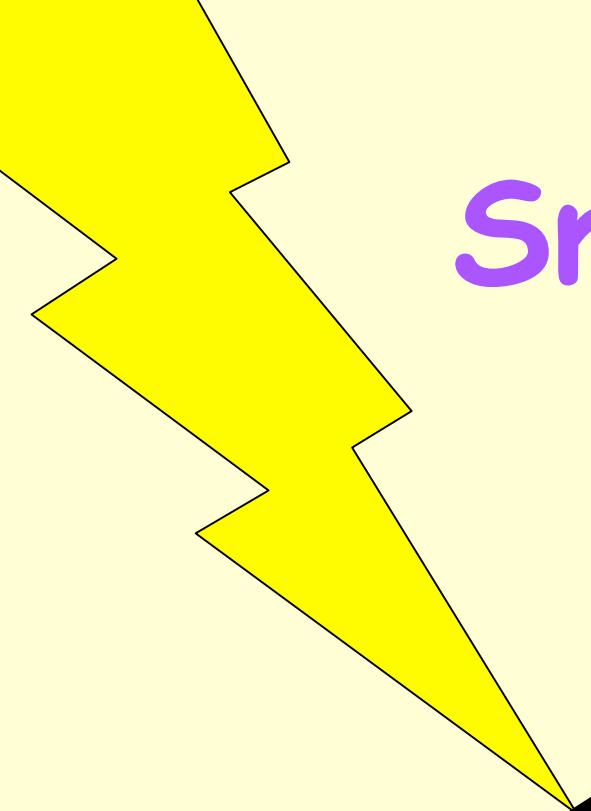
**Preservation
Action**

Long Term Storage



Thanks to Dave Tarrant
Southampton University

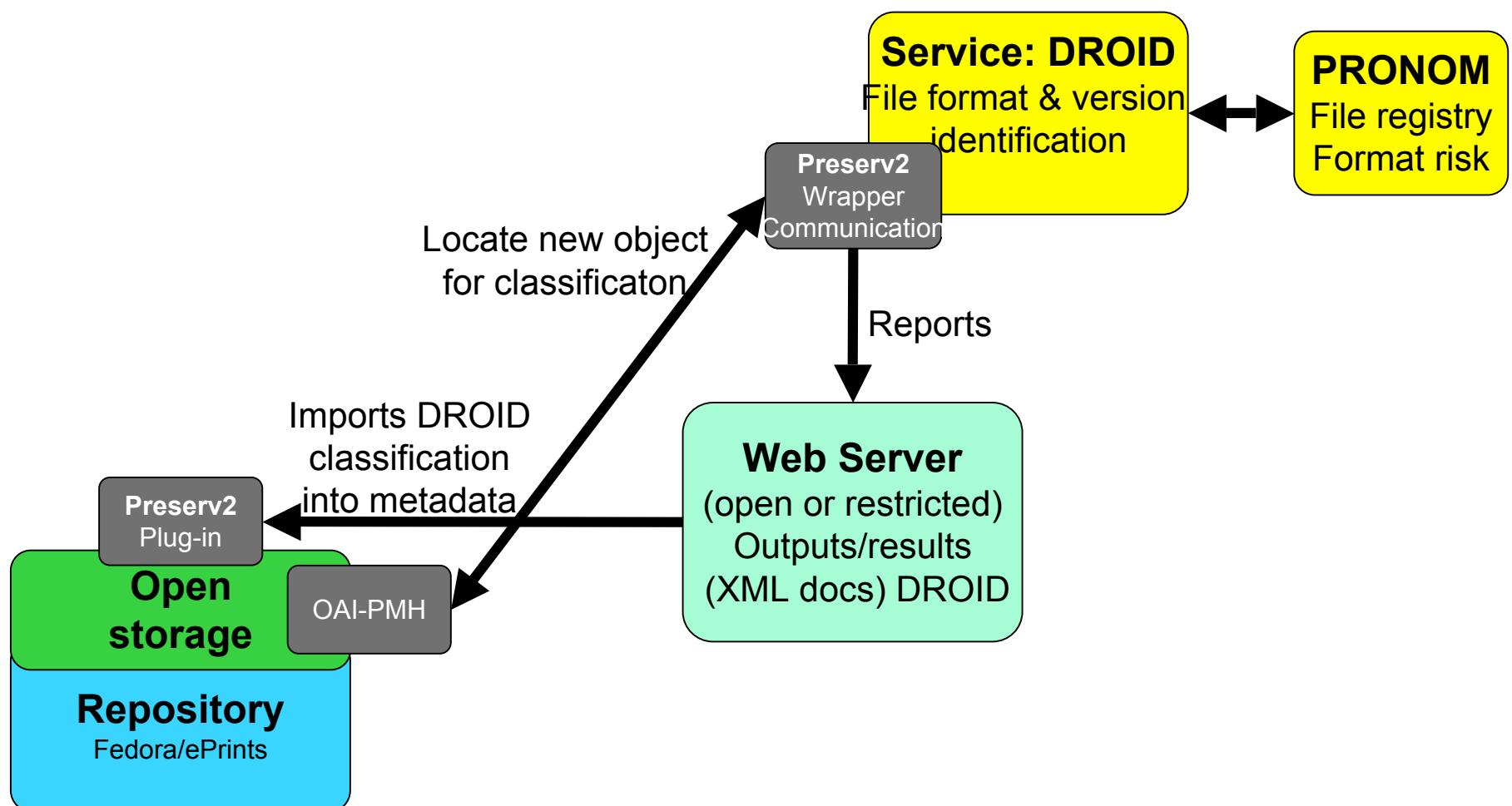


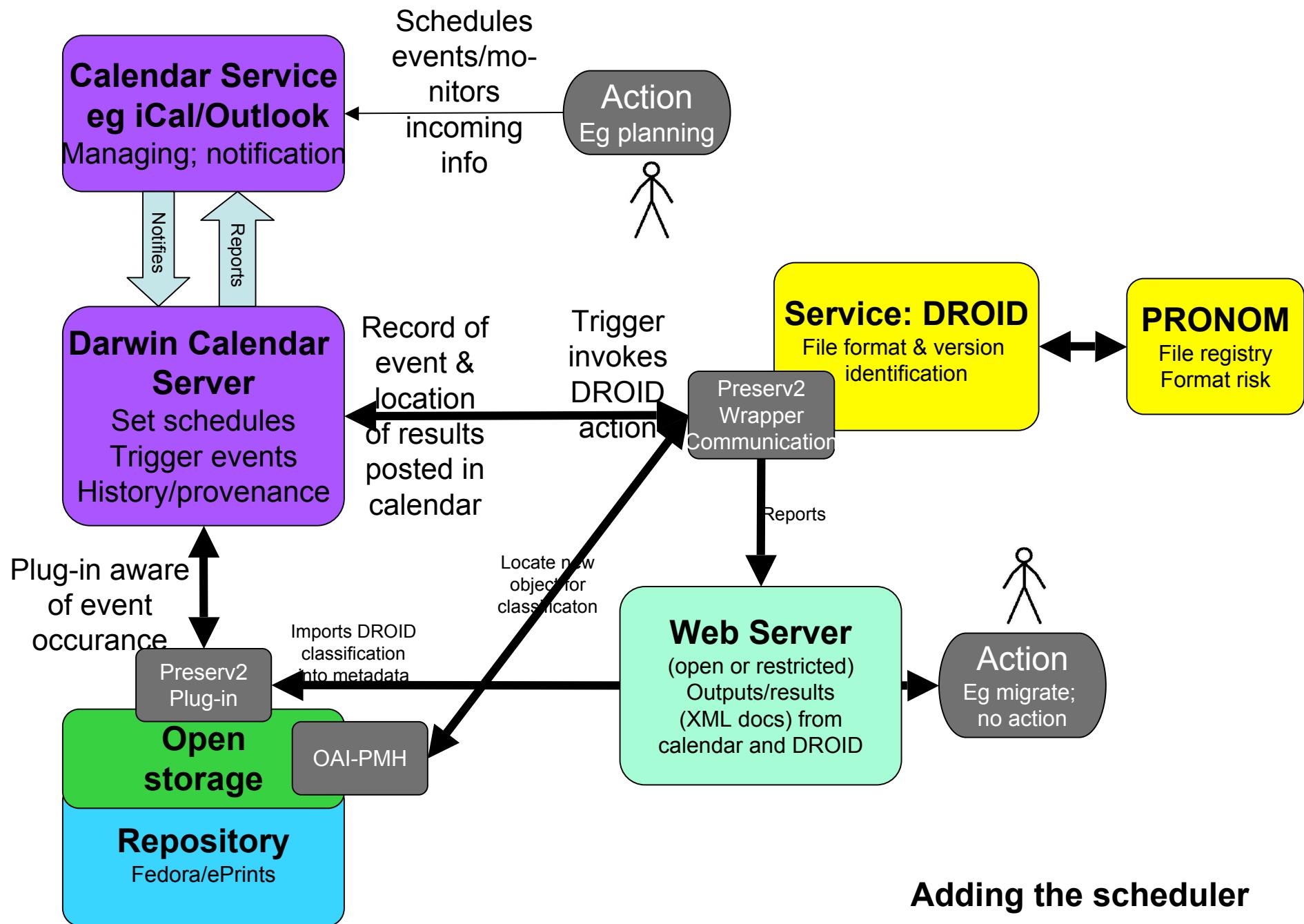


Smart storage

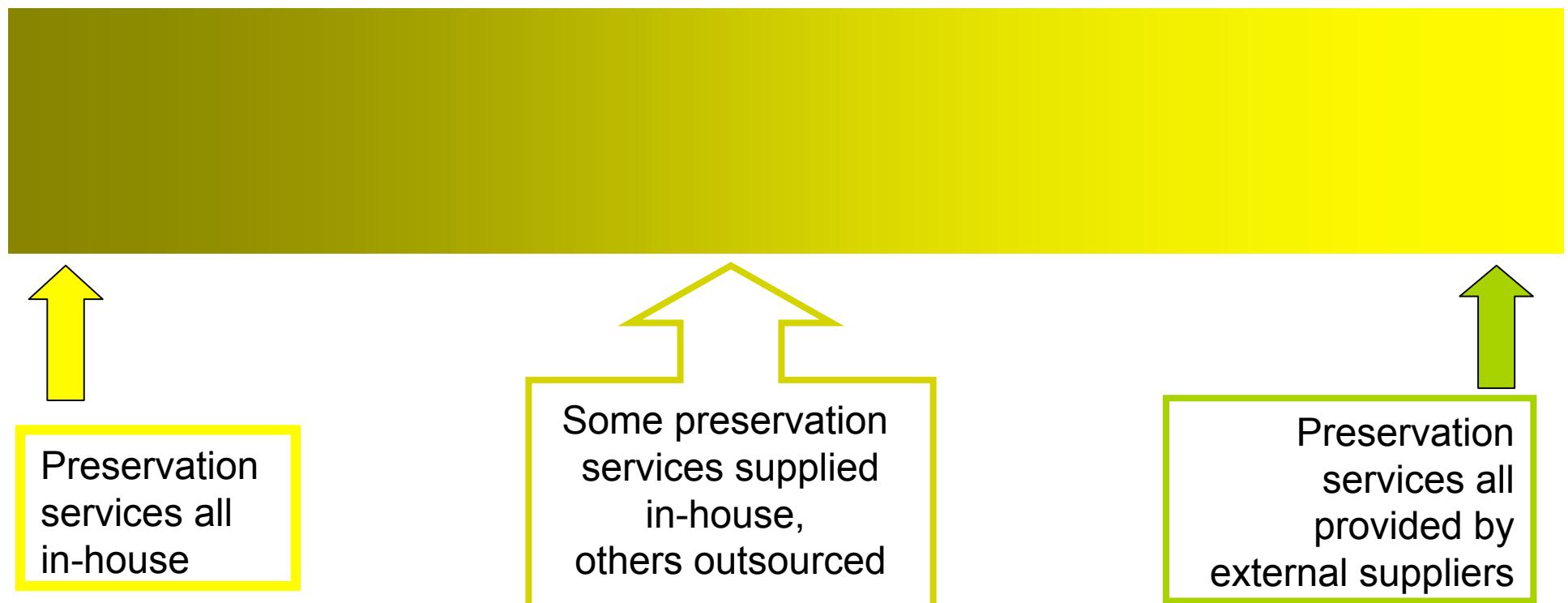
**combines an underlying
standard storage
approach with intelligence
provided through services**

Smart storage for preservation in action: Format identification and characterisation using DROID-PRONOM





Continuum of provision of repository preservation services





navigation

- Repositories Research
- About
- News
- Events
- Info and activities
- Projects
- Clusters
- Scenarios/use cases/UML
- Workflows
- Deposit API/SWORD
- SWAP
- CRIG
- FAQs
- Links
- Recent changes
- Sandbox
- Help

search

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version

[article](#) [discussion](#) [edit](#) [history](#)

CRIG

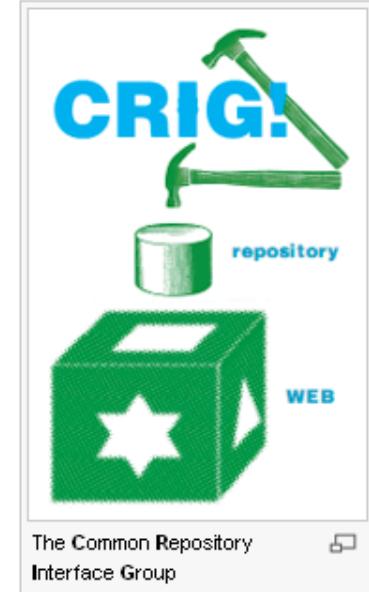
[[CRIG Home](#) | [CRIG Telcon 29/3/07](#) | [CRIG Scenarios](#) | [CrigCasts \(Podcasts\)](#) | [Unconference \(Bloomsbury, Dec 6/7\)](#) | [ReST Workshop \(Bloomsbury, Feb 8\)](#) | [RepoChallenge \(Southampton, April 2-4\)](#) | [DRY Barcamp \(Bath, Friday June 6\)](#) | [Roadshow \(July 08\)](#) | [RepoCamp \(LoC 25 July 08\)](#)]

Contents [hide]

- 1 JISC Common Repository Interfaces Group (CRIG)
- 2 Workplan
- 3 Membership
- 4 Events (Chronological Order)
- 5 Channels of Communication

JISC Common Repository Interfaces Group (CRIG) [edit]

“The coolest thing to do with your data will be thought of by someone else”



JISC set up the Common Repository Interfaces Group (CRIG) to help identify problem spaces in the repository landscape and suggest innovative solutions. The CRIG consists of a core group of technical, policy and development staff with repository interface expertise. It encourages anyone to join who is dedicated and passionate about surfacing scholarly content on the web. Please see below for joining details.

Workplan [edit]

For an interactive version of the diagram below (i.e. the image as a 'table of contents' that links to important pages throughout this wiki) please see: <http://www.flickr.com/photos/wocrig/2469877865/>

Define**Prototype****Establish**

Interoperability in Action



OR08 Publications

OR2008
Third International Conference
on Open Repositories

Conference Home | Repository Home | About | Browse by Session | Browse by People | Advanced Search | Login | Create Account

Welcome to OR08 Publications

This is a repository used to manage publications for the [Open Repositories 2008](#) conference. All eprint URLs are persistent and can be redirected to other repositories or services as required in the future. Hint: To download many items at once, perform a search and export the results in 'Zip' format.

- Main Conference
 - Keynote
 - Repository Developer Challenge Finalists
 - Tuesday 1st April
 - 2a - Social Networking
 - 2b - Sustainability (a)
 - 2c - Interoperability
 - Birds of a Feather (Tues)
 - Posters
 - Wednesday 2nd April
 - 4a - National Perspectives
 - 4b - Scientific Repositories (a)
 - 5a - Legal
 - 5b - Scientific Repositories (b)
 - 5a - Sustainability (b)
 - 5b - Models, Architectures & Frameworks
 - 5c - Tools
 - Birds of a Feather (Wed)
 - Repository Managers
 - Object Reuse and Exchange (ORE) Open Day

OR08 Publications supports [OAI-ORE](#) with a base URL of <http://pda.eprints.org.uk/ore/ore>



ora
OXFORD+UNIVERSITY
RESEARCH+ARCHIVE

ORA Basic Item: "Job Mobility of Residents and Migrants in Urban China"

Reference:
John Knight ; Linda Y. Yueh , (2003) Job Mobility of Residents and Migrants in Urban China .

Title : Job Mobility of Residents and Migrants in Urban China ;

Creator : John Knight ; Linda Y. Yueh ;

Date : 2003 ;

Subject : Classification-JEL-J21 ; Classification-JEL-J60 ; Classification-JEL-J63 ; Classification-EL-O53 ; labour mobility ; labour turnover ; layoffs ; China ; Application/pdf ;

Format : RePEc-handle:RePEc:oxf:wpaper:163

RePEc-number: 163

Original-URL: <http://www.economics.ox.ac.uk/Research/wp/pdf/paper163.pdf>

ora:1202

um:uuid:0ec1744f-e435-4176-86e9-f941003a8021

Browse | Contribute | Tools | Help

Downloads
[FULLTEXT](#) | [ASCII text of item](#) | [JOURNAL](#) | [Item](#)

Terms of Use
The copyright of this Item rests with the author and/or other copyright holders.
[Click here for our Terms of Use](#)

OAI-ORE
EPrints & Fedora

OR2008
Third International Conference
on Open Repositories

Search | Detailed Search | Browse | Tools

Preserv.org.uk

Repository Preservation and Interoperability

Browse:

Subject see more
2008-04 - (86)

Automatically derived terms see more

1 - Web 2.0 - (3)
2a - Social Networking - (3)
3 - Interoperability - (3)
4a - National Perspectives - (3)
4b - Scientific Repositories - (5)
5a - Legal - (3)
5b - Scientific Repositories - (5)
6a - Sustainability - (3)
6b - Models, Architectures & Frameworks - (3)

Format see more

application/pdf - (81)
application/vnd.ms-powerpoint - (10)
text/html - (2)

Creator see more

Carr, Leslie - (5)
Allison, Paul - (3)
Luhhard, Dill - (3)
Aver, Chris - (2)
Coles, Simon - (2)
Kahn, Jeffrey - (2)
Kumar, Anoop - (2)
Mandl, Peter - (2)
Namiki, Takao - (2)
Pepler, Sam - (2)

oxford
OXFORD+UNIVERSITY
RESEARCH+ARCHIVE

Job Mobility of Residents and Migrants in Urban China

, John Knight and , Linda Y. Yueh (2003) Job Mobility of Residents and Migrants in Urban China .

XML 1186b

JPG 955b

Plain Text 1362b

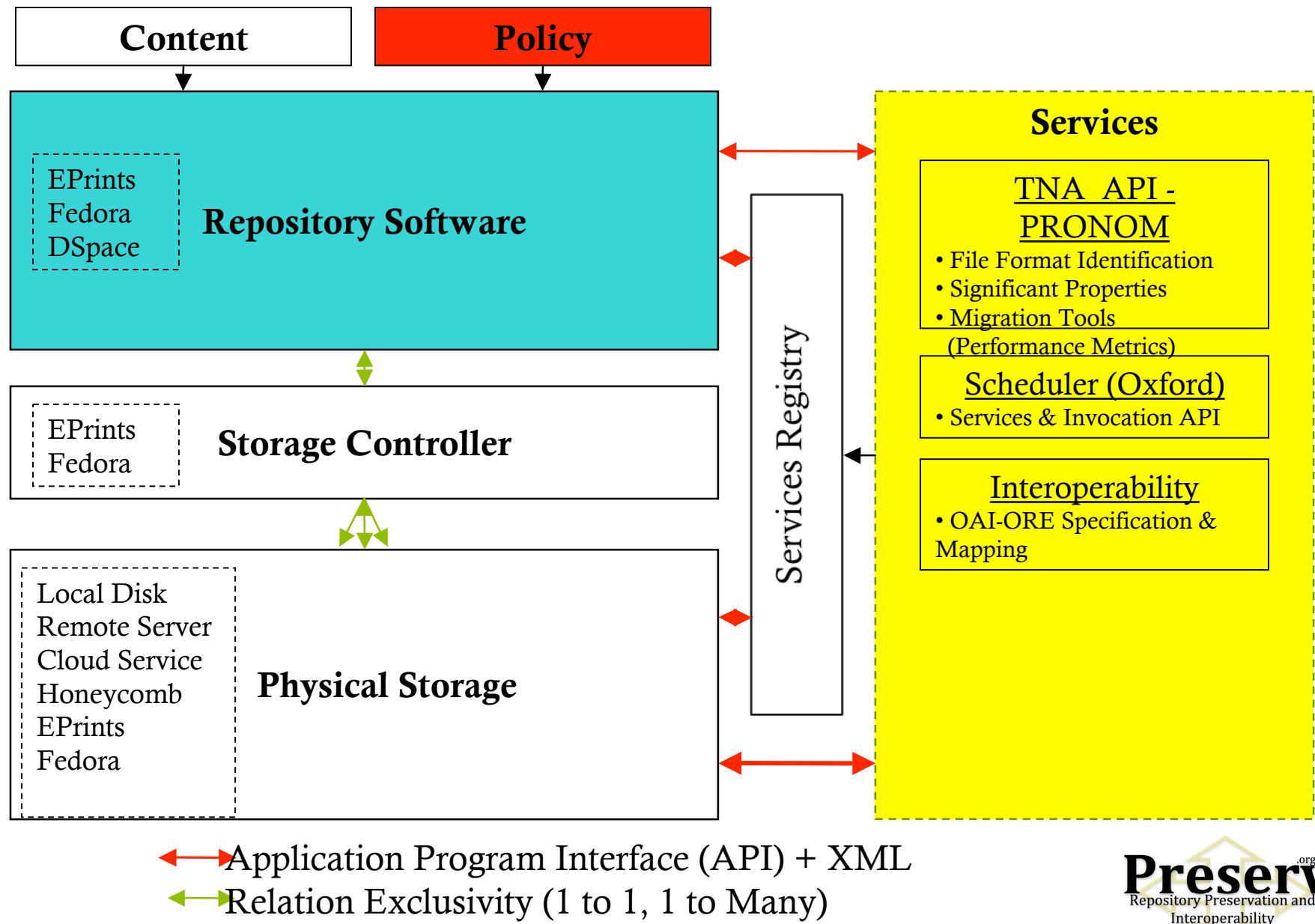
XML 678b

PDF - Requires a PDF viewer such as [GSview](#), [Xpdf](#) or [Adobe Acrobat Reader](#) 241Kb

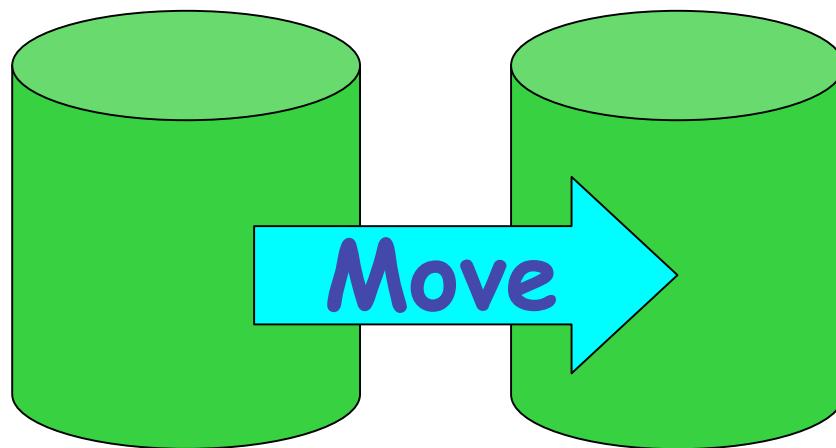
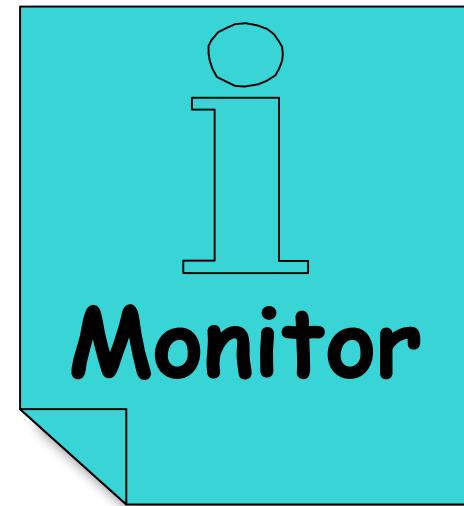
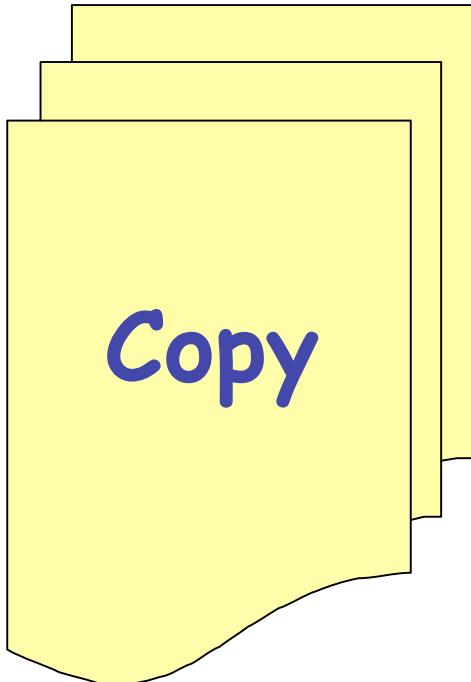
Plain Text 66Kb

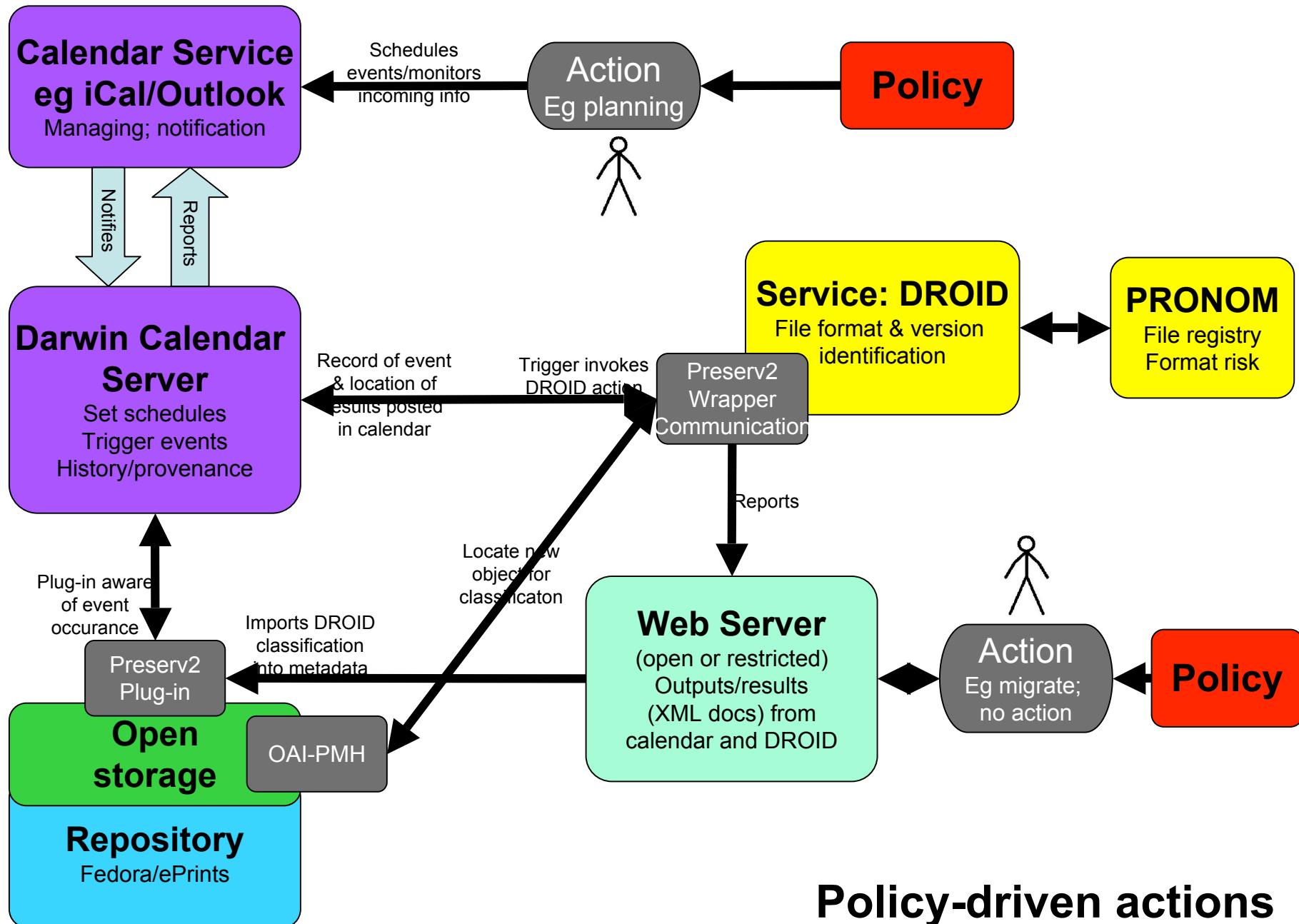
Which is which?

Preserv2 Project Structure



Three fundamental preservation actions on data





Formats/Risks



This EPrints install is referencing a trial version of the risk analysis service. None of the risk scores are likely to be accurate and thus should not be used as the basis for a program of action.

High Risk Objects

OLE2 Compound Document Format  1

[Hitchcock_a60_doc.doc](#) (278Kb)

Title: Towards smart storage for repository preservation services

EPrint ID: 4

User: Mr David C Tarrant

User	No of Files
Mr David C Tarrant	1

Medium Risk Objects

Microsoft Powerpoint Presentation (Version 97-2002)  3

Low Risk Objects

Portable Document Format (Version 1.4)  3

Portable Document Format (Version 1.3)  2

ZIP Format  2

PRESERV website

<http://preserv.eprints.org/>

Sally Rumsey sally.rumsey@ouls.ox.ac.uk

PRESERV Project Manager: Steve Hitchcock
sh94r@ecs.soton.ac.uk

