

The logo features the word "eprints" in a white, lowercase, sans-serif font. The letter "e" is enclosed within a white square frame that has a small blue triangle at its top-right corner. Below the word "eprints", the website address "www.eprints.org" is written in a smaller, white, lowercase, sans-serif font.

eprints
www.eprints.org

This is a faded, light blue version of the EPrints logo and website URL, identical in design to the one above but with significantly reduced opacity.

eprints
www.eprints.org

EPrints

Preservation Update



EPrints Preservation

- ▶ New EPrints Preservation Team



- ▶ Overseeing integration of preservation into EPrints software as well as involvement with preservation projects
- ▶ Preserv2 – Ended March 2009
- ▶ Keep-It! – Just Beginning



Keep-It! Project Overview

- ▶ Aim: To create a number of exemplar preservation repositories from which others can learn
- ▶ Small number of very diverse repositories



Training



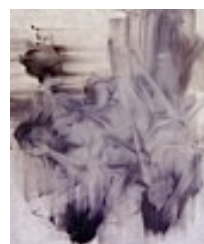
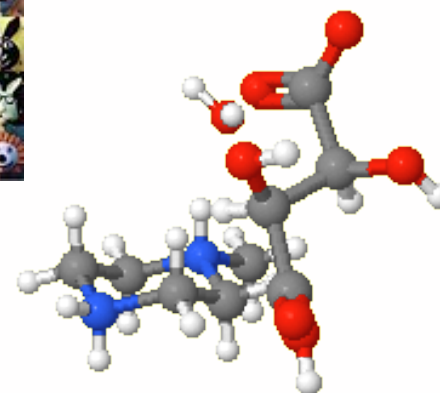
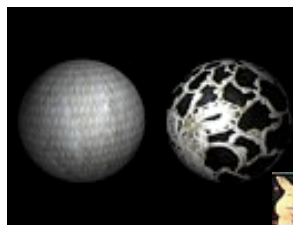
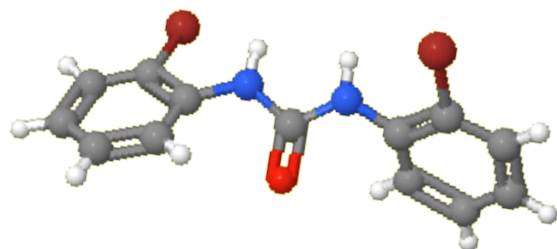
Development



Deployment

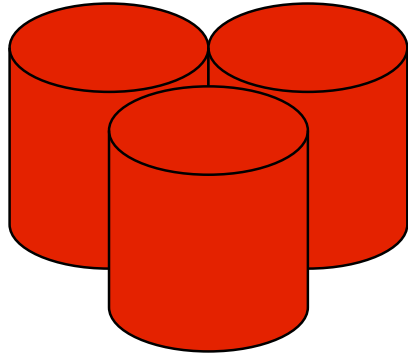


Repository Content





Keep-It Focuses



**Long Term
Reliable Storage**



Risk Analysis

Mitigation / Action





From Preserv2 - Risk Analysis



Preserv 2

eprints

Home | About | Browse by Year | Browse by Subject |


Logged in as Mr David C Tarrant | [Manage deposits](#) | [Profile](#) | [Saved searches](#) | [Review](#) | [Admin](#) | [Logout](#)

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


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

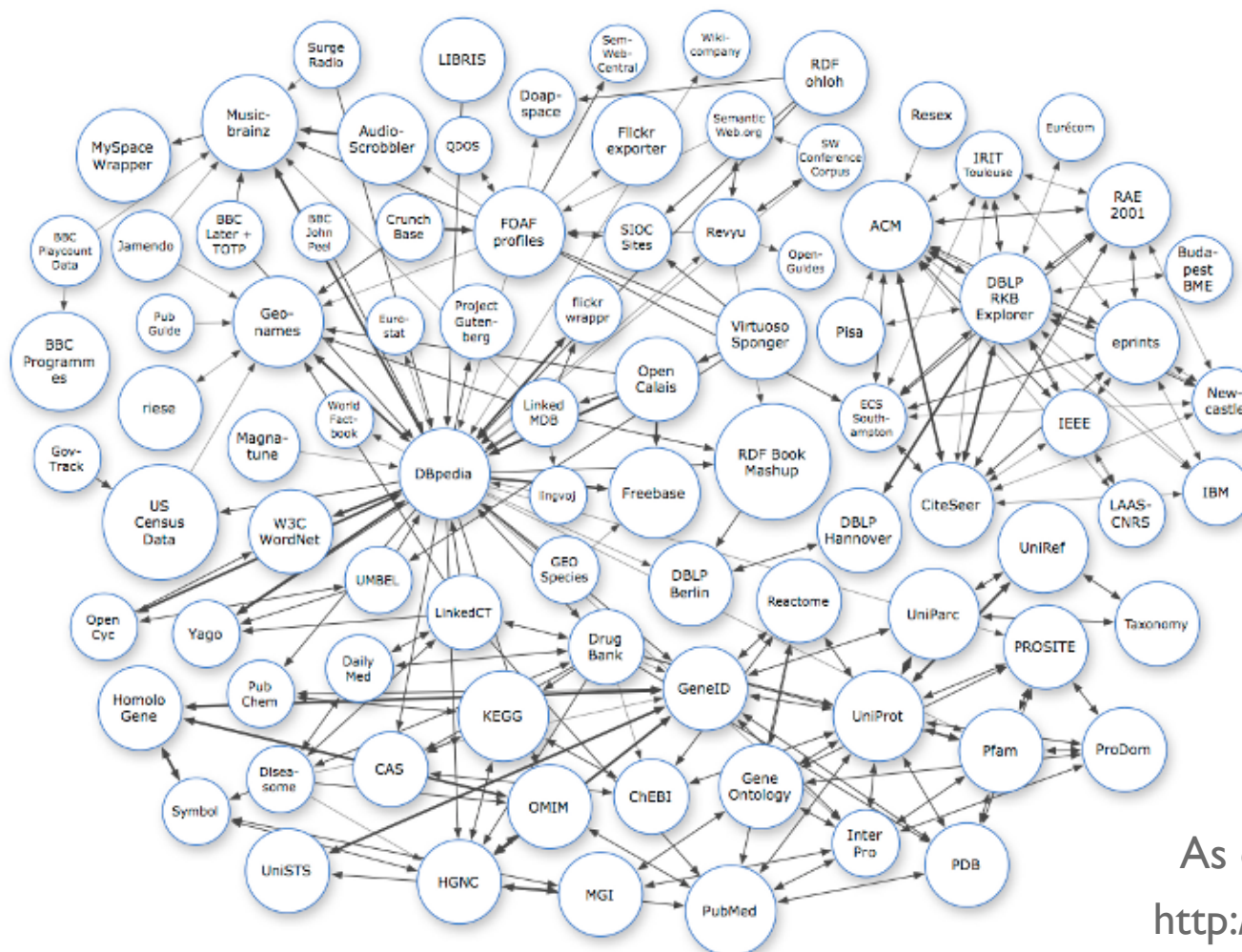


Risk Analysis - Problems

- ▶ Data pertaining to Risks is distributed across the web
- ▶ Bringing all this data together into one registry is hard if that registry has a well controlled data model
- ▶ Not so much of a problem if you harness linked data and the semantic web



Linked Data Web



As of March 2009
<http://linkeddata.org/>

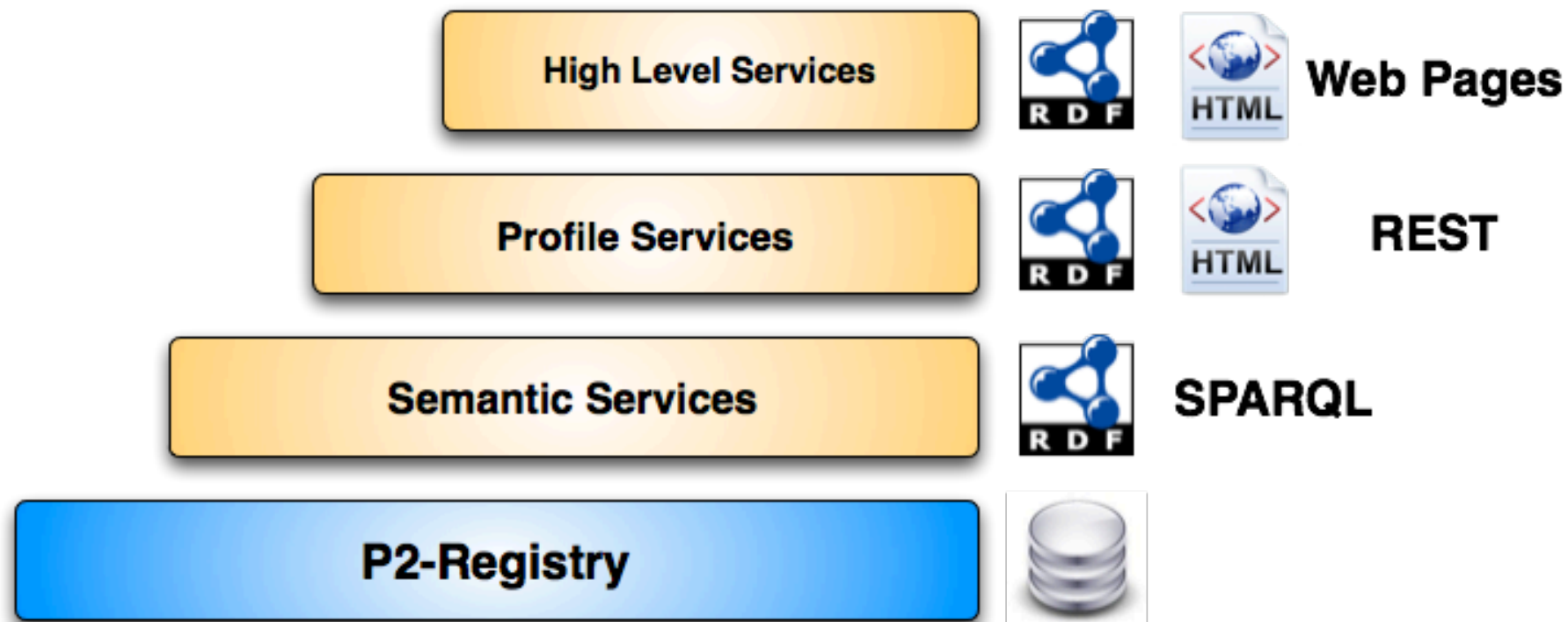


The P2-Registry

- ▶ A preservation focused registry
 - ▶ Imports openly available data (Pronom, dbpedia)
- ▶ Provides a set of restful services
 - ▶ Data export
 - ▶ Risk Analysis Profile Data
 - ▶ Processed Risk Analysis Data
 - ▶ Migration Pathways
- ▶ SPARQL Endpoint



The P2-Registry





Example Input – PDF v1.3

▲ PRONOM Data

- ▲ Format Documentation
- ▲ Format age and newer versions
- ▲ Identification Type
- ▲ 4 Software Tools

▲ dbpedia Data

- ▲ Format Complexity
- ▲ Format Type
- ▲ > 50 Software Tools



Example Output – PDF v1.3

Risk Analysis - Portable Document Format (v1.3) (Default Profile)

Portable Document Format (v1.3)

Format Age : Your format is 10 years old and there are 3 newer formats, the latest of which is PDF (1.6) (Released: 01 Jan 2004).

Software Tools (Open) : 3 tools can Open your format.

Software Tools (Save) : 1 tools can Save your format.

Format Documentation? : Documentation exists for this format

Documentation Quality : Documentation is complete and of a high standard

Rights : Format is proprietary

Portable Document Format

Format Age : Your format is 16 years old but is the latest known version of this format.

Ubiquity : Format is most widely adopted of type

Stability : Format is not backwards compatible, but versions change infrequently

Identification Type : Format can be positively identified (specific)

Format Type : It is not possible to obtain the original document in the original context using this format

Complexity : Medium complexity format

Software Tools (Open) : 14 tools can Open your format.

Software Tools (Save) : 39 tools can Save your format.

Risk Score: 3.73

Total = 41 / 11 properties

How is this calculated?

The data you see here has all come from the Preserv2 registry and more specifically the risk analysis service. Available [here](#) in RDF the risk analysis services selects specific information from the registry according to a profile (in this case the default one) and outputs in in RDF. This page displays a summary of this data which has also been process to find a score relating to this data.

Each piece of select data is either about the format itself or it's related supertype format, e.g. PDF 1.6 is a type of PDF. From this point data is handled in 4 ways with all final risk levels being either **low (green)**, **medium (orange)** or **high (red)**. To calculate the final risk score low risks are worth 1 point, medium - 5 points and high - 10 points. The total is then divided by the number of properties which counted towards this score to give the final risk score. Items with ~~lines~~ through them are not counted due to better or more accurate overriding information being available in a different category.

The risk boundaries are:

- <3.51 = Low Risk
- >3.50 and <7.00 = Medium Risk
- >=7.00 = High Risk



P2-Registry Futures

Import more data which people have created on the linked data web.

Currently 46,000 items of Information (mostly about PDF)

Have you got any?



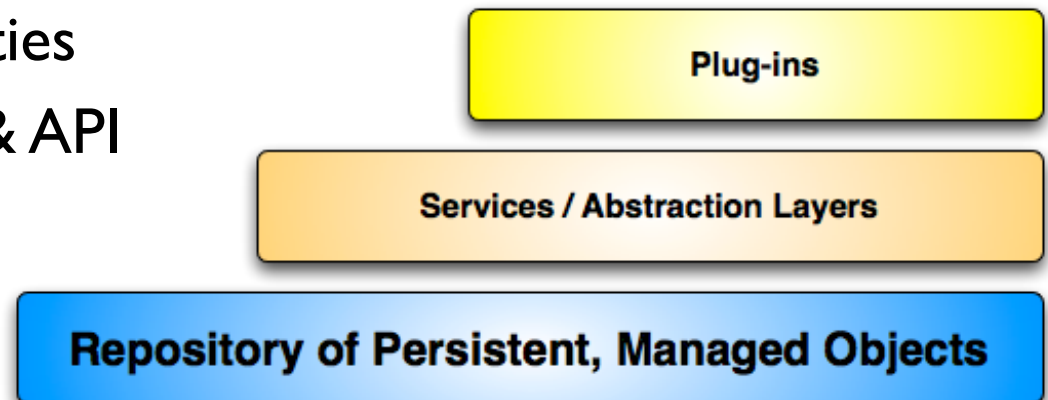
EPrints 3.2

Desktop & Cloud Integration



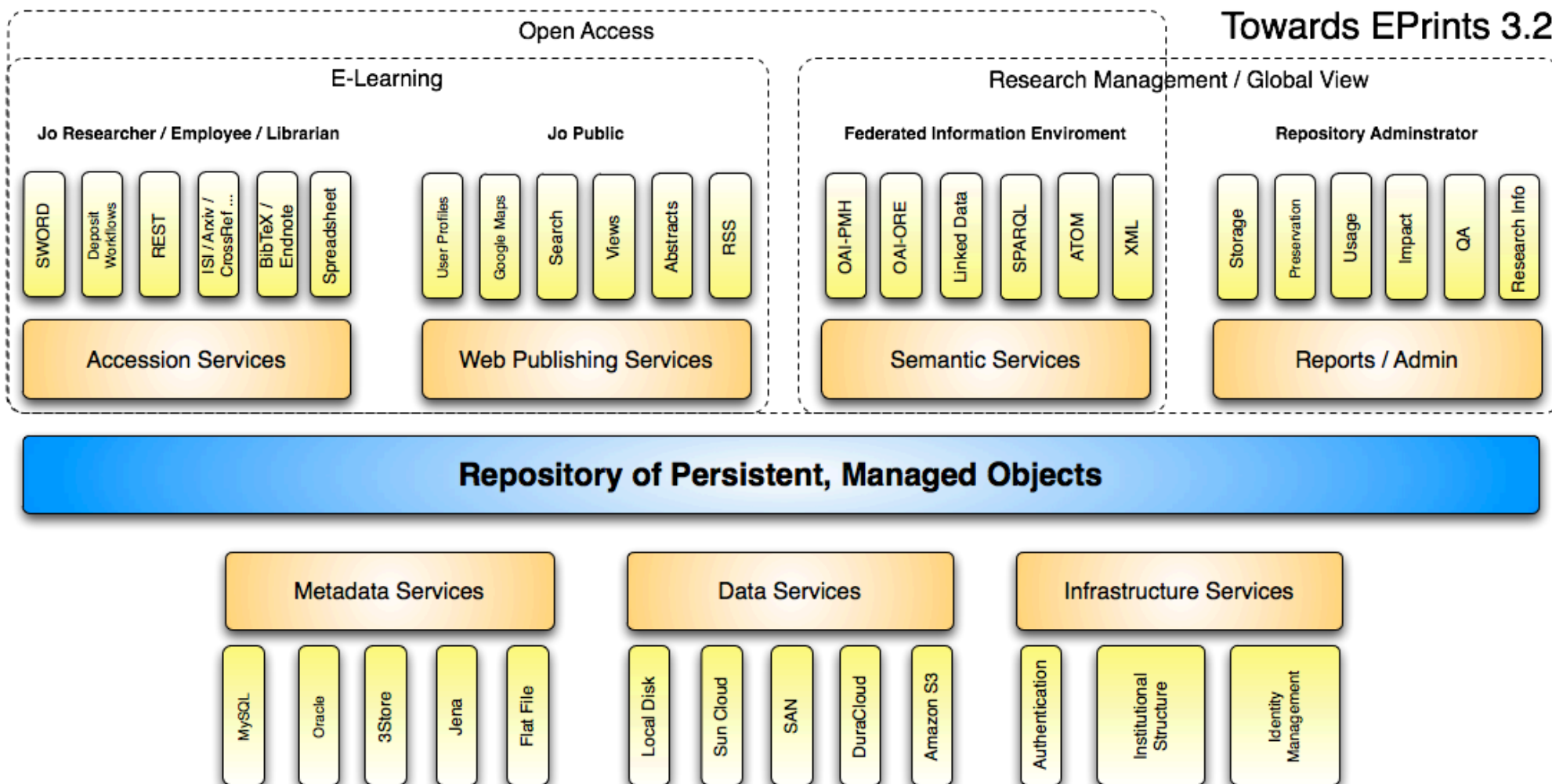
EPrints Structure

- ▶ Everything builds on the core layer
- ▶ Major part of v3.2 is strengthening the core and adding more abstraction layers
 - ▶ Improved data model
 - ▶ Enhanced data facilities
 - ▶ Enhanced metadata facilities
 - ▶ Improved programming & API



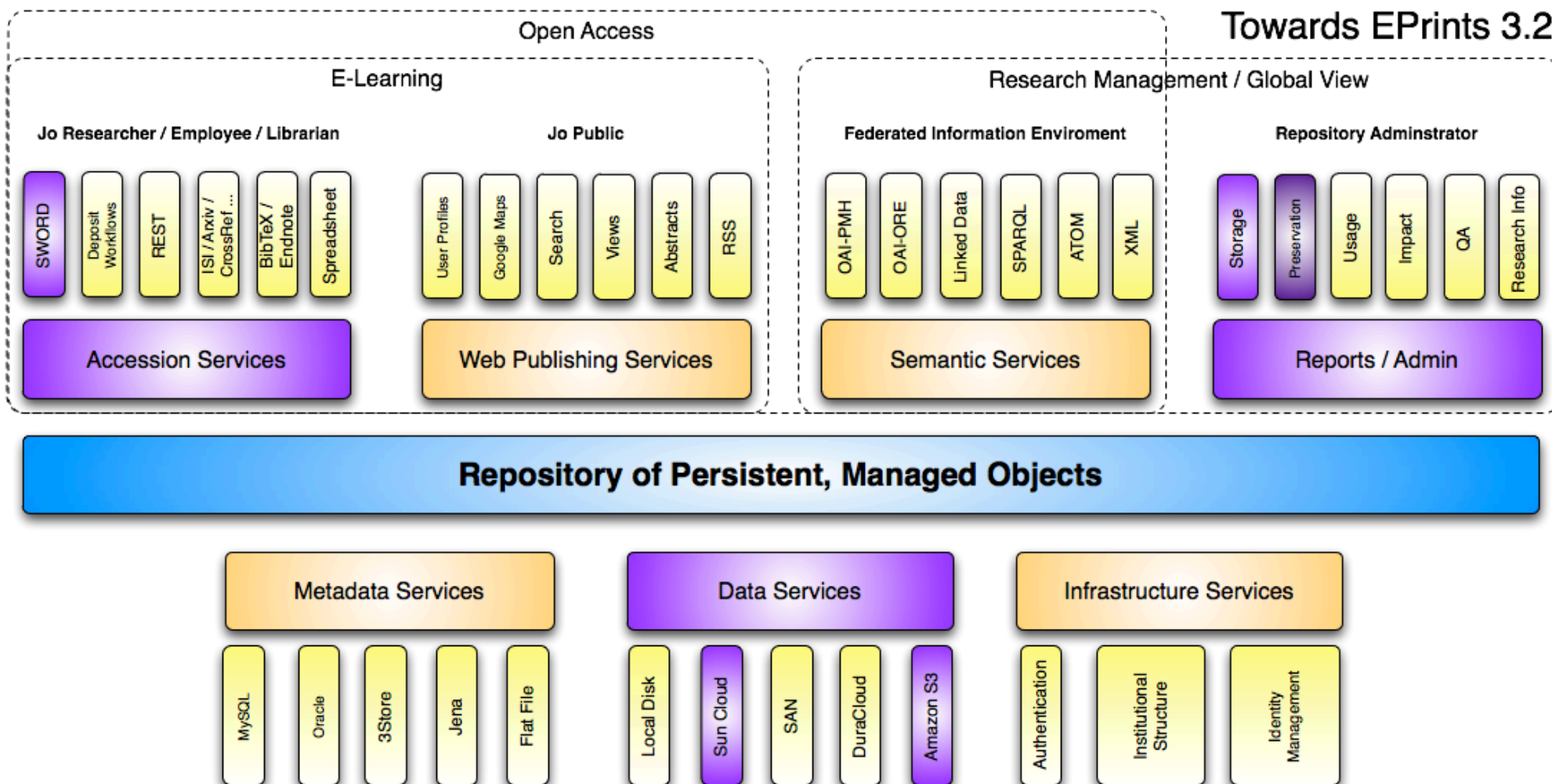


EPrints 3.2 Structure





Presentation Focus

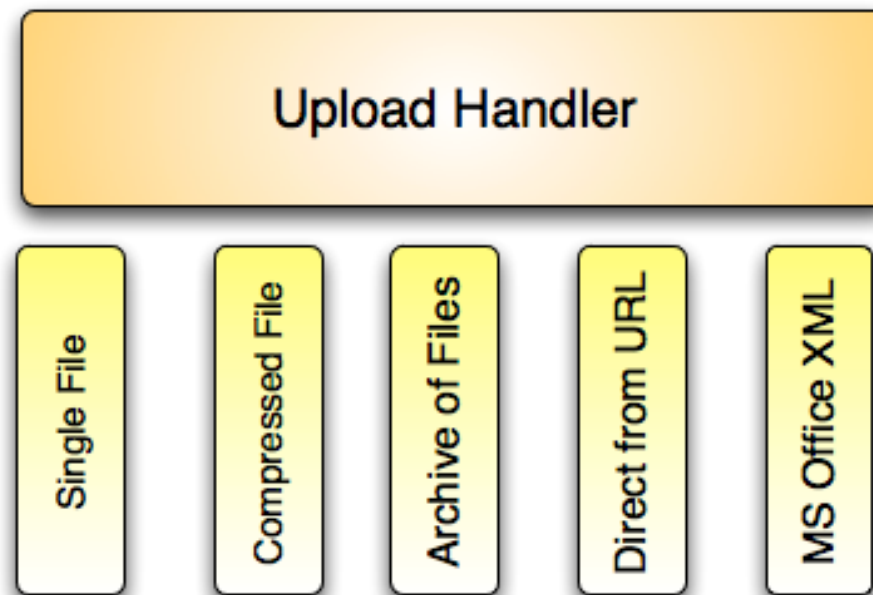




Desktop & Cloud Integration

Part I: Desktop

- Upload handler now has a Plug-in layer

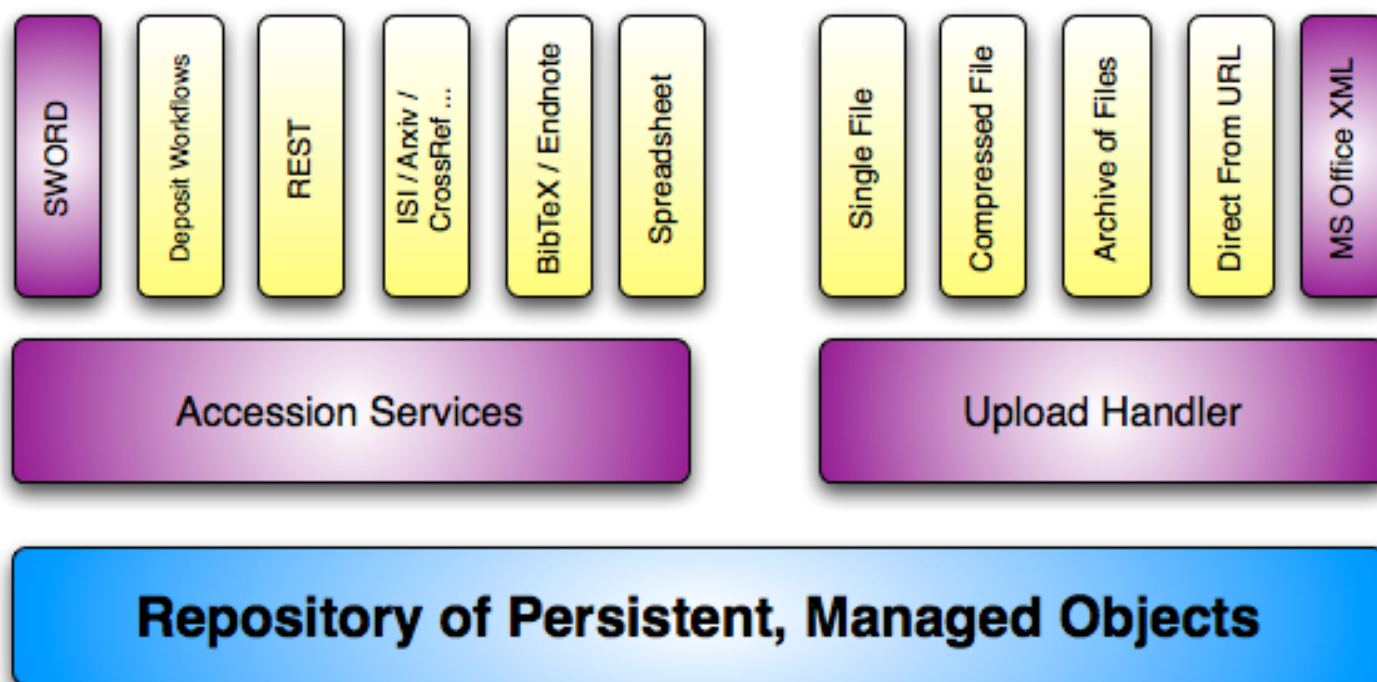




Desktop & Cloud Integration

Part I: Desktop

- Can be integrated with accession services





Desktop & Cloud Integration

Part I: Desktop

EPrints 3.2

- ▲ SWORD 2 Support
- ▲ MS Office XML Upload Plug-in

Microsoft Office

- ▲ Microsoft Article Authoring Add-in Tool
- ▲ Dublin Core Metadata
- ▲ SWORD 2 Support



Demo

Demo I

1) Uploading Office 2007 Document into EPrints using Web Interface

2) Uploading directly from Word 2007 into EPrints

Training 12

[Home](#) | [About](#) | [Browse by Year](#) | [Browse by Subject](#) | [Browse by Division](#)Logged in as Dr Leslie Carr | [Manage deposits](#) | [Profile](#) | [Saved searches](#) | [Review](#) | [Admin](#) | [Edit page](#) | [Logout](#) [Search](#)

Welcome to Training 12

Welcome to Training 12. [Click here to start customising this repository.](#) Atom  RSS 1.0  RSS 2.0

[Latest Additions](#)

View items added to the repository in the past week.

[Search Repository](#)

Search the repository using a full range of fields. Use the search field at the top of the page for a quick search.

[Browse Repository](#)

Browse the items in the repository by subject.

[About this Repository](#)

More information about this site.

[Repository Policies](#)

Policy for use of material in this repository.

Training 12 supports [OAI 2.0](#) with a base URL of <http://training12.eprints.org/cgi/oa12>

Manage deposits - Training 12 - Windows Internet Explorer


http://training12.eprints.org/cgi/users/home?screen=Items

Google

★ Favorites Manage deposits - Training 12

Home About Browse by Year Browse by Subject Browse by Division

Logged in as Dr Leslie Carr | [Manage deposits](#) | [Profile](#) | [Saved searches](#) | [Review](#) | [Admin](#) | [Edit page phrases](#) | [Logout](#)













Training 12 

Manage deposits

[? Help](#)


[New Item](#) [Import Items](#)

☒ User Workarea. ☒ Under Review. ☐ Live Archive. ☐ Retired.

Last Modified	Title	Item Type	Item Status	
18 Jun 2009 15:59	EPrints 3.2 - Preservation Slide Subset [PASIG Malta 09]	Article	User Workarea	   
30 May 2009 11:12	About me Website	Article	User Workarea	   
30 May 2009 10:55	Web Based File Uploading	Article	User Workarea	   

Abstract [Add Column](#)

Training 12 is powered by [EPrints 3](#) which is developed by the [School of Electronics and Computer Science](#) at the University of Southampton. [More information and software credits.](#)



Internet | Protected Mode: On 100%

16:20 18/06/2009



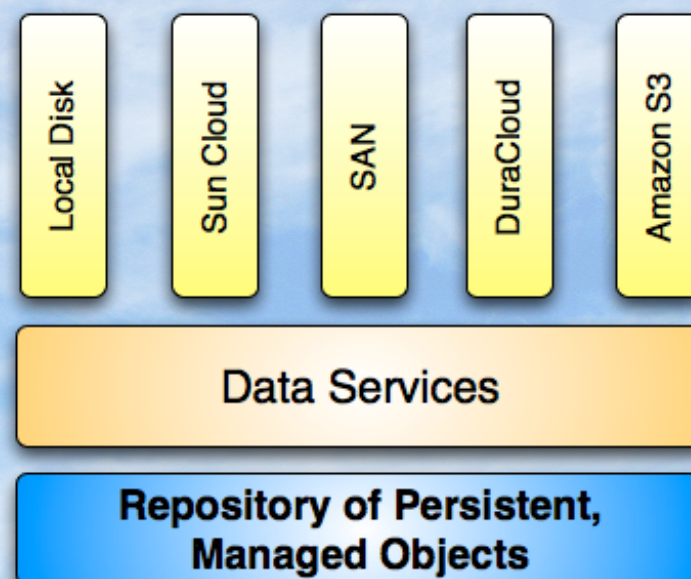
Desktop & Cloud Integration

Part II: Cloud – Hybrid Storage

- Using a single storage platform or solution has drawbacks.

Cost vs. Speed vs. Reliability

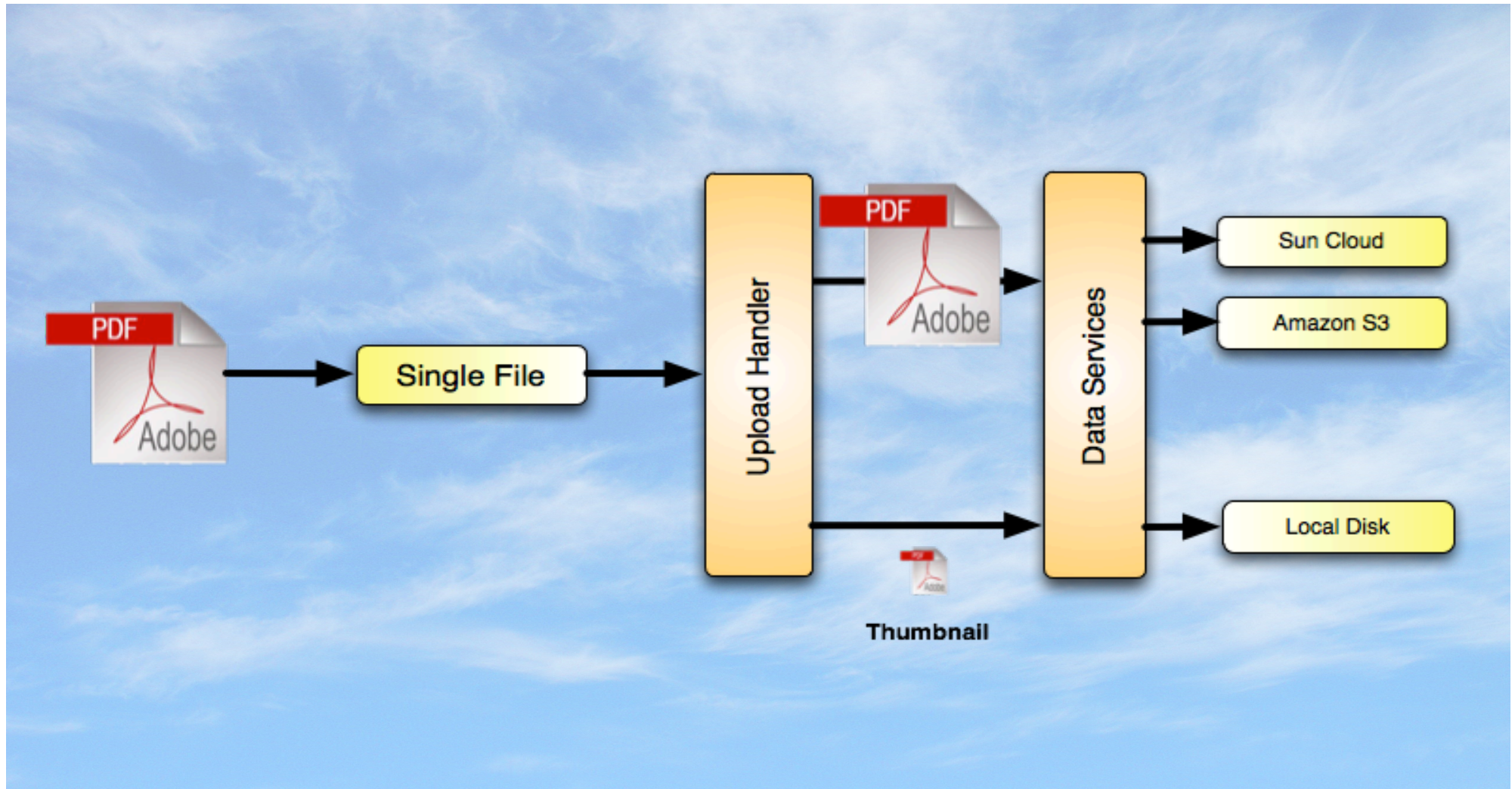
- If repositories are to provide good preservation then they need to utilize and be able to migrate to new platforms.
- In this presentation we look at backing your repository with a Hybrid storage solution.
- Gives you the power to utilize the benefits of each solution without losing control of your digital objects.





Desktop & Cloud Integration

Part II: Hybrid Storage Policies





Desktop & Cloud Integration

Part II: Hybrid Storage Policies

```
<choose>  
  <when test="datasetid = 'document'">  
    <choose>  
      <when test="$parent{relation_type} = 'http://eprints.org/relation/isVolatileVersionOf'">  
        <plugin name="Local"/>  
      </when>  
      <otherwise>  
        <plugin name="SunCSS"/>  
        <plugin name="AmazonS3"/>  
      </otherwise>  
    </choose>  
  </when>  
  <otherwise>  
    <plugin name="Local"/>  
  </otherwise>  
</choose>
```




Demo

Demo II

I) Uploading a PDF into the cloud

I b) Managing your storage

Training 12



[Home](#) [About](#) [Browse by Year](#) [Browse by Subject](#) [Browse by Division](#)

Logged in as Dr Leslie Carr | [Manage deposits](#) | [Profile](#) | [Saved searches](#) | [Review](#) | [Admin](#) | [Edit page phrases](#) | [Logout](#)

Search

Manage deposits

[? Help](#)

[New Item](#)

[Import Items](#)

☒ User Workarea. ☒ Under Review. ☐ Live Archive. ☐ Retired.

Last Modified	Title	Item Type	Item Status	
18 Jun 2009 16:48	From the Desktop to the Cloud: Leveraging Hybrid Storage Architectures in your Repository	Article	User Workarea	
18 Jun 2009 16:44	EPrints 3.2 - Preservation Slide Subset [PASIG Malta 09]	Article	User Workarea	
30 May 2009 11:12	About me Website	Article	User Workarea	
30 May 2009 10:55	Web Based File Uploading	Article	User Workarea	

Abstract

Add Column



Desktop & Cloud Integration Summary

- ▶ Direct from desktop upload methods:
 - ▶ MS Word 2007 + Article Add-in via SWORD
 - ▶ FTP
 - ▶ WebDav – Drag and drop into your repository

- ▶ Storage Controller
 - ▶ Local Plug-ins (Local Disk, Local Compressed)
 - ▶ Local Archival Plug-ins (Honeycomb, NAS, SAN)
 - ▶ Cloud Plug-ins (Amazon S3, SunCSS, Flickr)



Demo

Demo III

Direct Upload From Microsoft Word 2007 to the Cloud

Admin - Training 12 - Windows Internet Explorer

http://training12.eprints.org/cgi/users/home?screen=Admin

Google

Home About Browse by Year Browse by Subject Browse by Division

Logged in as Dr Leslie Carr | [Manage deposits](#) | [Profile](#) | [Saved searches](#) | [Review](#) | [Admin](#) | [Edit page phrases](#) | [Logout](#)

Search

Training 12

eprints

Admin

- Editorial Tools
- System Tools
 - [Storage Manager](#) - Manage your repositories hybrid storage setup, including migration of files.
 - [Reload Configuration](#) - Cause the repository to re-read its configuration files. Use with caution!
 - [View Configuration](#) - View configuration files for this repository.
 - [Phrase Editor](#) - Modify phrases used in the system.
 - [Edit subject](#) - Modify the subject tree(s)
 - [Manage Metadata Fields](#) - Add or remove metadata fields
 - [Check Subjects](#) - Check subject tree integrity
- Config. Tools
- Misc. Tools

Training 12 is powered by [EPrints 3](#) which is developed by the [School of Electronics and Computer Science](#) at the University of Southampton. [More information and software credits.](#)

eprints

Internet | Protected Mode: On

100%

17:11
18/06/2009



Thank-You

Questions?