

DIRECTIONS FOR EPRINTS

Dr Leslie Carr
University of
Southampton,
Web Science
Institute
& EPrints
Services



<http://www.flickr.com/photos/coderkind/408839789/>

Decades of Achievement

Internet (packets)



Web (documents)



Repository (preservation)



What have we achieved with this multilayered platform in the last decade?
What scope have we for achieving new things?



EPrints History

- Conception
 - Oct 1999
- Birth
 - June 2000
- Version 2
 - Feb 2002
- Version 3
 - Jan 2007

EPrints Timeline (1994 – 2010)

NSF - JISC Open Citations Project
October 1999 – December 2002

OpCit project

The project, funded by the joint NSF-JISC International Network of Research Programmes, brought together Steven Hamill (who had recently joined JCS) with Les Carr and Steve Hitchcock to investigate citation linking in the academic literature.

Scite: a hyperlink each of the over 100,000 pages in Los Alamos' unique online Physics Archive to every other paper in the archive that it cites.

Context: a system to integrate a family of generic linking tools and to design author and user interfaces to enable easy adoption by other archives.

Links: a system that provides a remarkable new way of navigating the scientific journal literature and induce authors in other fields to create interlinked online archives like Los Alamos across disciplines and around the world.

Registry of Open Access Repositories (ROAR)
January 2004

Created initially to track the growth of EPrints repositories, and quickly expanded to deal with repositories based on any software.

JISC ECrystals data repository
June 2004

Developed by the EdShare JISC project. Ecrystals is the archive for Crystal Structures generated by the Southampton Chemical Crystallography Group and the PRSC UK National Crystallography Service.

Berlin J. Open Access
March 2005

Progress in Implementing the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities.

JISC IRRA (Institutional Repositories and Research Assessment project)
June 2005

Research Assessment is a complex activity involving decisions made by many parties (individual researchers, institutions, funding bodies, peer review, journal assessment) and potentially involving many interoperating software systems. This project aims to develop practical solutions to the problems of data exchange between and repository workflows into RAE activities (both from institutional and national perspectives) and to advise on how to deploy these solutions in their local contexts.

JISC Interoperable Repository Statistics
June 2005 – May 2007

A JISC project investigating and developing a joint approach to OA statistics gathering and sharing.

JISC Repository for the Laboratory project
June 2005 – May 2007

A research project investigating and developing digital data and document repositories for laboratory-based sciences.

EPrints Services launched
September 2005

EPrints Services offers a complete range of commercial services to support EPrints, including hosting, training and consultancy.

JISC Repository Support Project
October 2006 – March 2009

Led by U Nottingham, the aim of RSP is to co-ordinate and deliver good practice and practical advice to English and Welsh institutions on the development, management and development of digital institutional repositories.

PrInS v3 released
January 2007

EPrints version 3.0 released at OR2007, and described as 'a significant milestone toward a repository software' according to a review by Peter Wilkinson and William Nixon in *Arabian*, issue 50, January 2007.

JISC MyExperiment
March 2007 – March 2009

The myExperiment Virtual Research Environment enables scientists to share digital assets associated with their research projects, facilitate the sharing and execution of scientific workflows.

JISC Institutional Data Management Blueprint
October 2009 – March 2010

Creating a practical and sustainable institutional framework for managing research data that facilitates ambitious national and international e-research practice.

JISC Citation Services project
January 2010 – December 2010

This project (led by JCS) forms Strand 5 of the ongoing JSC Repository Infrastructure activity.

JISC Open Impact
January 2010 – October 2010

Collect evidence about the impact of research that has been undertaken by universities and to present it to a range of stakeholders (government, industry, the public, and third independent third party agency (a learned society). The project will be delivered through a series of workshops mediated through a particular society (the British Computer Society).

Subversive Proposal
June 1994

The "Subversive Proposal" was an Internet posting by Steven Hamill on 27.11.1994 (presented orally at the 1994 Network Services Conference in London on November 28). It proposed that the Internet be used as a vehicle for "subversive" research, not for "income" – to archive "theses" for all online (in anonymous FTP archives or web-sites) and to make them available to many of which were collected and published as a book in 1995: "Subversive Journals at the Crossroads: A Subversive Proposal for Electronic Publishing".

This led to the creation in 1997 of Cogprints, an Open Access Archive and affiliated archive in the Cognitive Sciences at the University of Southampton. It was the Scientific Open Access Forum (initially called the "Subversive Forum" until 1998) that became the Budapest Open Access Initiative, which coined the term "Open Access". The GNU Eprints software for creating OA repositories and the Cogprints software for creating OA repositories also evolved out of the Subversive Proposal.

The "Subversive Proposal" was in many ways ahead of its time, and summarized in the American Scientist Open Access Forum (http://www.americanscientist.org). The "Subversive Proposal" at 10: A retrospective was also written by Richard Poynder: "Ten Years After" Information Today 21 (9) October 2004

EPrints announced
October 1999

The first activity of the OpCit project was to fund Steven Hamill and Les Carr to attend the UPS meeting in Santa Fe, subsequently renamed the Open Archiving Initiative (OAI). At the end of the two-day meeting, Steven Hamill, Steven Hamill and Les Carr agreed that Southampton would develop a Turnkey system to help groups start up their own archives. This became EPrints.

JISC ELIB Open Journals Project
May 1995 – May 1998

A project that built on the concept of open hypermedia (as exemplified by previous Micromedia and DLS projects) to create a framework for publishing that allowed journals on the Web to be interlinked and that would reuse the existing infrastructure of the Internet. The outcome of the Subversive Proposal at 10: A retrospective was also written by Richard Poynder: "Ten Years After" Information Today 21 (9) October 2004

JISC ELIB Cogprints Project
January 1996 – December 1997

An ELIB Phase 2 program in "Cogprints" (the grey logo) that aimed to produce an analogue to seaserv.soton.ac.uk, a subject repository for the cognitive sciences. Project director was Steven Hamill, then from the Department of Psychology at the University of Southampton.

American Scientist Open Access Forum
September 1998

Forum devoted to the freeing of online access to the peer-reviewed research literature.

EPrints Software Released
June 2000

EPrints v1.0 released at the first official OAi workshop.

Cogprints converts to EPrints
January 2001

EPrints v2.0 released

EPrints v2.0 released
February 2002

OSI EPrints Handbook
June 2003

JISC EBank UK (phase 2)
February 2005 – January 2006

EPrints used to develop a Contextualized Learning Activity Repository.

Etheses mandate
September 2008

JISC KULTUR
March 2007 – March 2009

Kultur is JISC-funded project which is creating a model of an institutional repository for use in the creative and applied arts.

JISC VAURIC
April 2003 – September 2008

Increase trust in the scholarly content of repositories by validating the author postprint version of research articles against the publisher's preprint version available from the publishers' official web site.

JISC EdSpace
October 2007 – March 2009

To support sharing and collaboration in learning and teaching, EdSpace aims to change culture in the University of Southampton by facilitating all involved in education to work together to share and reuse resources. This is achieved by implementing social "Web 2.0" style tools for a repository based on EPrints. Culturally, this will be achieved by encouraging staff and students to think about how they will best benefit from such change.

ICrystals Federation
November 2007 – January 2009

This project will establish a solid foundation of crystallography data repositories across an international group of institutions, and increase the number of aggregator services. The project is led by the UK National Crystallography Service (University of Southampton) with cross-institutional partners from the Rutherford Appleton Laboratory, the University of Bath, the University of Edinburgh and the University of Cambridge. The project consists of 14 supporting partners.

ESRC NCRM repository
November 2007

The ESRC National Centre for Research Methods (NCRM) aims to improve the standards of research methods across the social sciences. The NCRM repository is a central access point to all NCRM outputs: books, articles, working papers, reviews, reports, presentations etc.

Open Repositories Conference 2008
March 2008

Repositories play a pivotal role in the evolving scholarly communication landscape, as a means of curating research data and scholarly collections. With its theme of "Practice and Innovation", OR2008 will create an opportunity for practitioners and researchers to share experiences and to explore the challenges of the new scholarly communication.

Master thesis repository
June 2008

Captures the dissertations and reports of Masters' students projects in Southampton. Not exposed off-campus.

JISC Knigf
April 2009 – September 2010

A JISC project to enable a diverse range of digital content presented by institutional repositories - research papers, datasets, images, video and audio - to be managed effectively today, tomorrow and beyond.

JISC Radiants & RIF
April 2009 – March 2011

Investigate and implement how to streamline the RIF data gathering exercise by building an interoperable institutional infrastructure, including repositories, that would capture and manage research outputs and other data.

ESRC CAHO journal
July 2009

Repository for 'Centre for Archaeology of Human Origins' journal

JISC IP200 project
June 2010 – November 2010

This project enables the EPrints repository to support the submission of XML-formatted experimental data together with the manuscript to which they correspond, and vice versa. The project is led by the University of Southampton. The basic motivation for the work are to promote the conservation of experimental data and to link data to publications.

JISC G4AC
June 2010

develop a prototype demonstrator that synthesizes research information from heterogeneous sources (institutional repositories and research council information systems), integrates the data and provides a user interface, and presents it to research-focused end users through an interface that will allow them to explore the state of the research landscape in UKRI.

dotAC.info

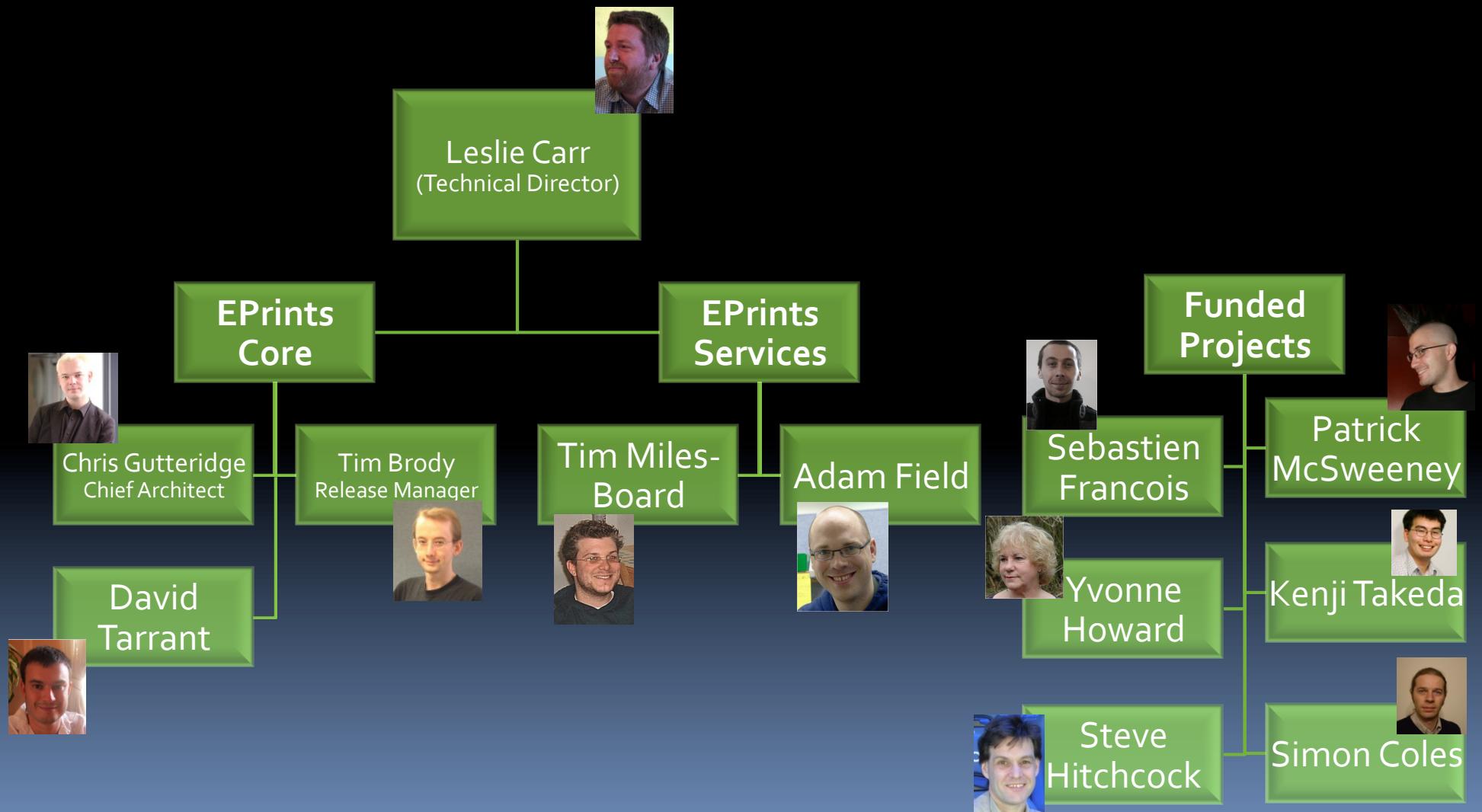
JISC Institutional Data Management Blueprint
October 2009 – March 2010

Create a practical and sustainable institutional framework for managing research data that facilitates ambitious national and international e-research practice.

JISC Citation Services project
January 2010 – December 2010

This project (led by JCS) forms Strand 5 of the ongoing JSC Repository Infrastructure activity.

EPrints Org Chart



EPrints Services

- For-pay commercial hosting, bespoke development, training, support
 - two fulltime repository developers/supporters
- Provides resources to sustain EPrints O/S development
- Provides customer contact
- Contributes code back to O/S
- Funds a PhD studentship
 - Bibliographic & social network analysis

What is the Point?

- Universities and researchers are knowledge producers and knowledge consumers
- The Web has radically altered the potential for knowledge dissemination in society in ways that we haven't yet fathomed
- We want to understand and facilitate that change
 - Research *and* development

Openness Agendas

- Open Access to Research Outputs
 - Supported by all Research Councils
 - Heavily promoted by JISC
- Open Educational Resources
 - Move from “Learning Objects” to “Learning Resources” via Open Access practice and methodology
 - Move from closed VLEs to sharing resources
 - All underpinned by active preservation and curation policies and workflows
- Open Research Data
 - Move from closed VRE / EScience environment
 - To open (able) data

Open Access to Research Outputs

- “A new public good”

Proteomic analysis of cell models of ovarian cancer tumour suppression
Sinclair, J.R. (2010) Proteomic analysis of cell models of ovarian cancer tumour suppression. Doctoral thesis, UCL (University College London)

PDF - Requires a PDF viewer such as [GSview](#), [Xpdf](#) or [Adobe Acrobat Reader](#)
5Mb

Excel Spreadsheet (Appendix 1A: whole cell lysate TMT)
4Mb

Excel Spreadsheet (Appendix 1B: secreted TMT)
1434Kb

Excel Spreadsheet (Appendix 1C: membrane TMT)
2056Kb

Abstract

Epithelial ovarian cancer (EOC) is the most common form of gynaecological malignancy in the developed world. Identifying molecular markers of disease may provide novel approaches to screening and could enable targeted treatment and the design of novel therapies. In previous work, 141 primary ovarian tumours were analysed using metaphase comparative genomic hybridization to identify complete or partial chromosome deletions that may harbour tumour and/or metastasis suppressor genes. Chromosome 18 (Ch18) was found to have deletions in 50% of the tumours. Microcell-mediated chromosome transfer (MMCT) of normal Ch18 material into EOC cell lines resulted in hybrids that displayed significant suppression of anchorage-independent growth. Tumourigenicity was reduced in a dose-dependent manner. In

UCL LIBRARY SERVICES
UCL Eprints

Library home » Online resources » UCL Eprints

Home :: Search :: Browse :: Information :: Statistics :: Latest additions :: Contact

Welcome to UCL Eprints

Quick search

Enter search term(s) Search

Information and help

[More](#) about UCL Eprints including FAQ, copyright, deposit guide and research funders' policies.

UCL's Open Access Policy

UCL Academic Board has approved an Open Access policy, ensuring that, subject to copyright permissions, all UCL research is placed online in UCL Eprints, freely available to all. For more details, see the June 2009 [Press Release](#). A UCL Publications Board has been convened to oversee the implementation of the UCL Open Access Policy.

For further information about UCL Eprints and Open Access generally, see [About UCL Eprints](#).

Research Funders' policies on Open Access

Holders of Wellcome Trust awards should consult the UCL Library Services [Guide to Wellcome Trust open access funding](#) for advice on compliance with the Trust's grant conditions. For advice on compliance with the policies of other research funders, please visit the Library's [Research Funder Policies](#) pages.

Statistics - most downloaded

Annual	2009	2008
Quarterly	Last quarter	All
Monthly	Last month	All

June 2010

Eprint	Downloads
Blow, L. and Leicester, A. and Oldfield, Z. (2003) <i>London's congestion charge. IFS Briefing Notes</i> (BN31). Institute for Fiscal Studies, London, UK.	1219
Sinclair, J.R. (2010) <i>Proteomic analysis of cell models of ovarian cancer tumour suppression</i> . Doctoral thesis, UCL (University College London)	1097
Owor, M. (2010) <i>Groundwater - surface water interactions on deeply weathered surfaces of low relief in the Upper Nile Basin of Uganda</i> . Doctoral thesis, UCL (University College London)	453
Krashia, P. (2009) <i>Recombinant glycine receptors: stoichiometry and kinetics</i> . Doctoral thesis, UCL (University College London)	294
Dustmann, C. and Casanova, M. and Fertig, M. and Preston, I. and Schmidt, C.M. (2003) <i>The</i>	

[Del.icio.us](#) [reddit](#) [StumbleUpon](#) [Facebook](#) [Digg](#)

Open Arts

- JISC KULTUR
consortium of
universities
and art
colleges

- Capturing performance, practice, exhibition, evidence
 - In a *no publication* environment

Capturing artistic experimentation & performance vs capturing scientific experimentation

Open Educational Resources

- JISC EdSpace
- IEEE LOM metadata vs Web 2, informally tagged
- Sufficient for discoverability and usefulness

The screenshot shows a digital repository interface for 'EdShare' at the 'UNIVERSITY OF Southampton'. The top navigation bar includes 'Home', 'Browse', 'About', 'Help', a search bar, and a 'Search' button. Below the navigation is a menu bar with 'Profile', 'Share manager', 'Bookmarks', 'Saved searches', 'Logout', and a 'Logout' button. The main content area displays a poster titled 'Posters for Web Science DTC Industrial Day'. The poster is titled 'What are the actual effects of Enterprise 2.0?' and includes sections on 'Origins of social software', 'Enterprise 2.0 Typology', 'Research methods', 'The hype and the question', and a 'Theoretical framework'. It is attributed to 'Mark Schueler (mcs08@ecs.soton.ac.uk)'. Below the poster are download links for 'Mark Schueler poster.pptx' (Microsoft PowerPoint Download), 'Sarosh Khan poster.ppt' (Microsoft PowerPoint Download), 'Michael Yip poster.pdf' (PDF Download), and 'Mike Santer.pptx' (Microsoft PowerPoint Download). To the right of the poster is a 'Description' box stating that the posters were created by Web Science MSc and PhD students as a discussion point with representatives from the DTC Industrial Advisory Group. Below this is an 'Advice for reuse' box stating that the posters are shared to facilitate ongoing discussions between students, the DTC faculty, and Industry. Further down is a 'Share details' box with information about the poster's creation, including 'Added By: Dr Leslie Carr', 'Added On: 16 Feb 2010 13:37', 'Creators: Leslie Carr, Catherine Pope', 'Tags: Web science', 'Viewing permissions: World', 'Link: http://www.edshare.soton.ac.uk/4790/Downloads:', and a 'Downloads' section with a small line graph. At the bottom is a 'Toolbox' with links for 'Edit item', 'Bookmark this item', 'Download as zip', and 'Use as template'.

Open Scientific Data

The screenshot shows a web page for the eCrystals database. The title is "(6-Thien-2-ylpyrid-3-yl)methanol". The page includes the following information:

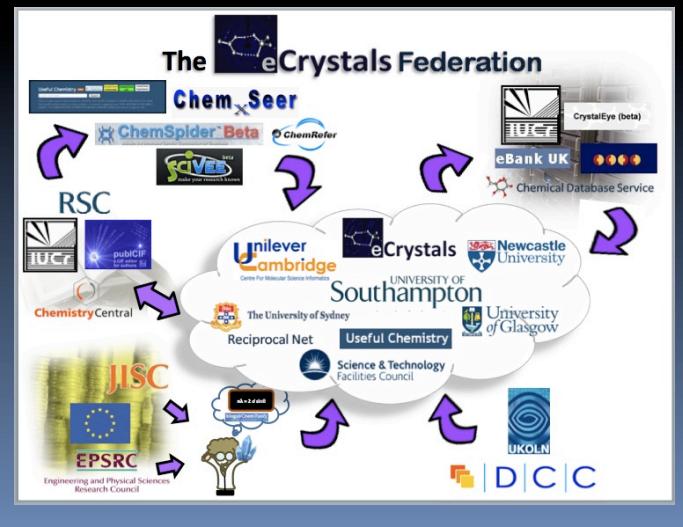
- Sample Originator: Donated Sample^a.
- Data Collection: Susanne L. Huth^b and Michael B. Hursthouse^b
- Structure Determination: Susanne L. Huth^b.
- Sample kindly provided by Maybridge^a
University of Southampton^b
- Chemical formula: C₁₀H₉NOS
- InChI: InChI=1/C10H9NOS/c12-7-8-3-4-9(11-6-8)10-2-1-5-13-10/h1-6,12H,7H2
- Identification Number: 10.3737/ecrystals.chem.soton.ac.uk/373
- Date Created: 09 July 2007
- Deposited On: 31 Jan 2008 13:08
- Deposited By: Ms L.S. Huth
- Depositor Comments: More information about this compound can be found at: <http://www.maybridge.com/>
- Data collection parameters:

Chemical formula	C ₁₀ H ₉ NOS
Crystallisation Solvent	
Crystal morphology	Needle
Crystal system	Orthorhombic
- Available Files:
 - Final Result: 2007may0014.cif (11k), 2007may0014.cml (3k), 2007may0014.fcf (99k)
 - Validation: 2007may0014_checkcif.htm (8k)
 - Refinement: 2007may0014.res (4k), 2007may0014_xl.lst (30k)

A repository of chemical data. Scientific, not bibliographic, metadata.

JISC EBank project (2003)

Run by the UK Crystallographic Service and latterly an international consortium.



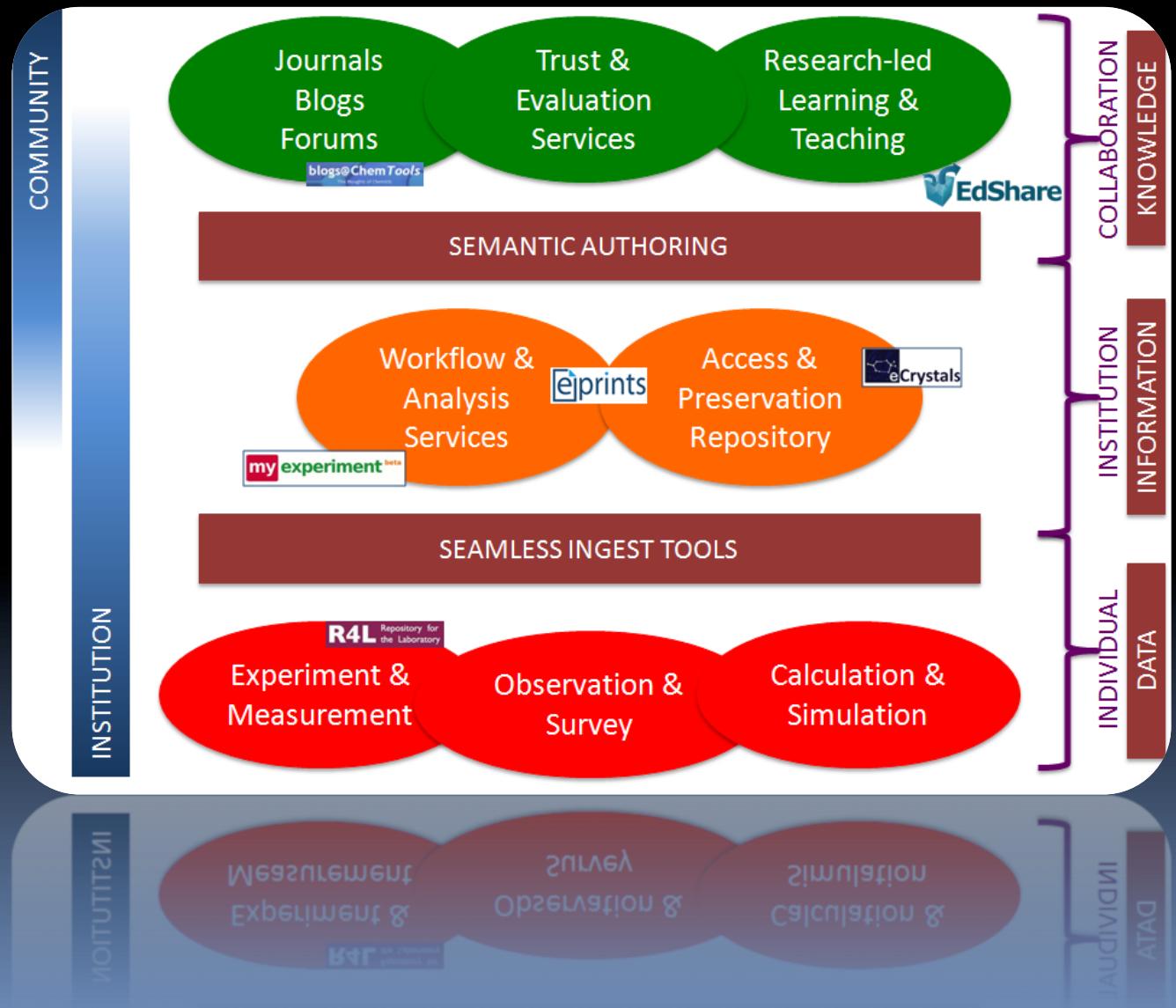
(Other Open Research)

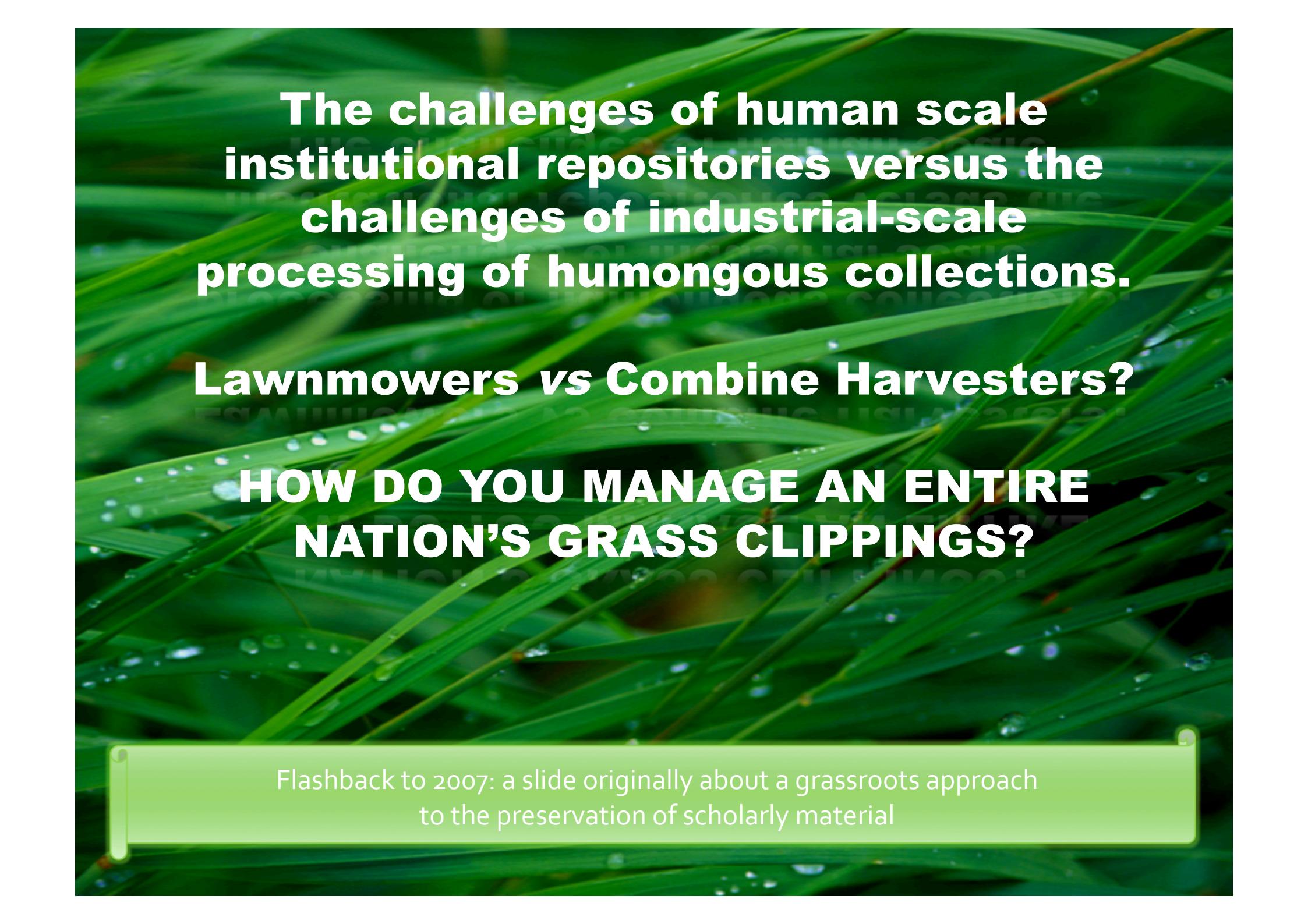
- JISC myExperiment
- Capturing
 - Workflows
 - Data
- Comes from EScience/ Cyberinfrastructure
- But uses Web2 as design principle

The screenshot shows the myExperiment website interface. At the top, there is a navigation bar with links for 'About', 'Mailing List', 'Publications', 'Log In', 'Register', 'Give us feedback', and 'Invite'. The 'Workflows' tab is currently selected. The main content area displays a workflow entry for 'Lymphoma type prediction based on microarray data'. The entry includes details such as 'Version 7 (latest) (of 7)', 'Workflow Type: Taverna 2', 'Original Uploader: Wei Tan', 'License: CC BY SA', 'Credits: (3)', 'Attributions: (0)', and 'Tags: (11)'. On the right side, there is a sidebar for 'Log in / Register' with fields for 'Username or Email', 'Password', 'Remember me', and 'Use OpenID'. Below the sidebar, there are links for 'Need an account? Click here to register' and 'Forgot Password?'. The 'Popular Tags' section lists various tags like 'benchmarks', 'bioinformatics', 'blast', 'grid', 'data integration', 'ebi', 'example', 'graph', 'gwas', 'gworkflow', 'kegg', 'localworker', 'microarray', 'mygrid', 'onDEX', 'ovtk', 'oxl', 'pathway', 'pathways', 'protein', 'pubmed', 'rdf', 'taverna', and 'text mining'. The 'Tags (11)' section lists 'Original Uploader tags' including 'cabig', 'cagrid', 'cancer', 'research', 'classification', 'gene', 'expression', 'globus', 'grid', 'service', 'lymphoma', 'lymphoma', 'workflow', 'microarray', 'support vector machine', and 'Log In to add Tags'. The 'Shared with Groups (0)' section is currently empty.

Understanding Data Management

- JISC Institutional Data Management Blueprint (IDMB)
- Campus-wide study to produce a 10-year plan





**The challenges of human scale
institutional repositories versus the
challenges of industrial-scale
processing of humongous collections.**

Lawnmowers vs Combine Harvesters?

**HOW DO YOU MANAGE AN ENTIRE
NATION'S GRASS CLIPPINGS?**

Flashback to 2007: a slide originally about a grassroots approach
to the preservation of scholarly material

EPrints Activity

Research Business

Research Management

Research Processes

Research Impact

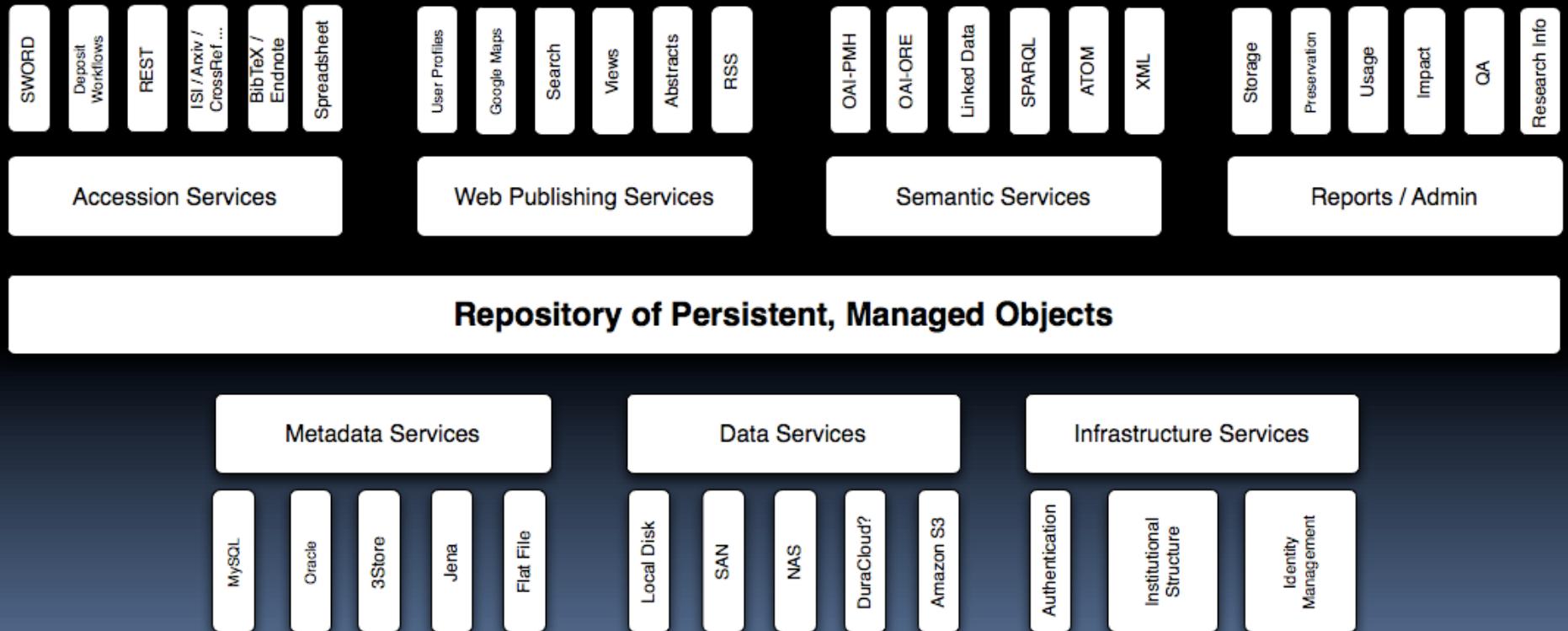
Research Content

Research Publications

Educational Resources

Scientific Data

EPrints Software Architecture

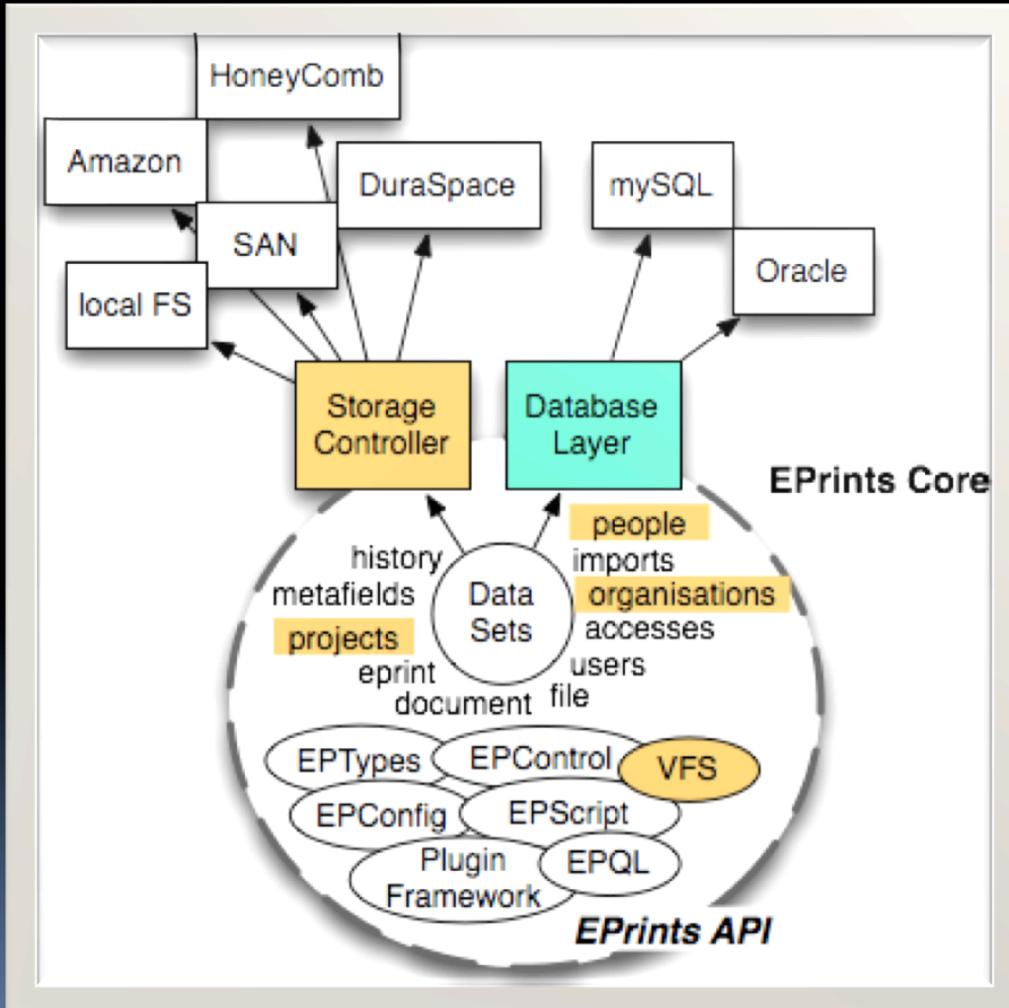


EPrints Core

Introduced in 3.2

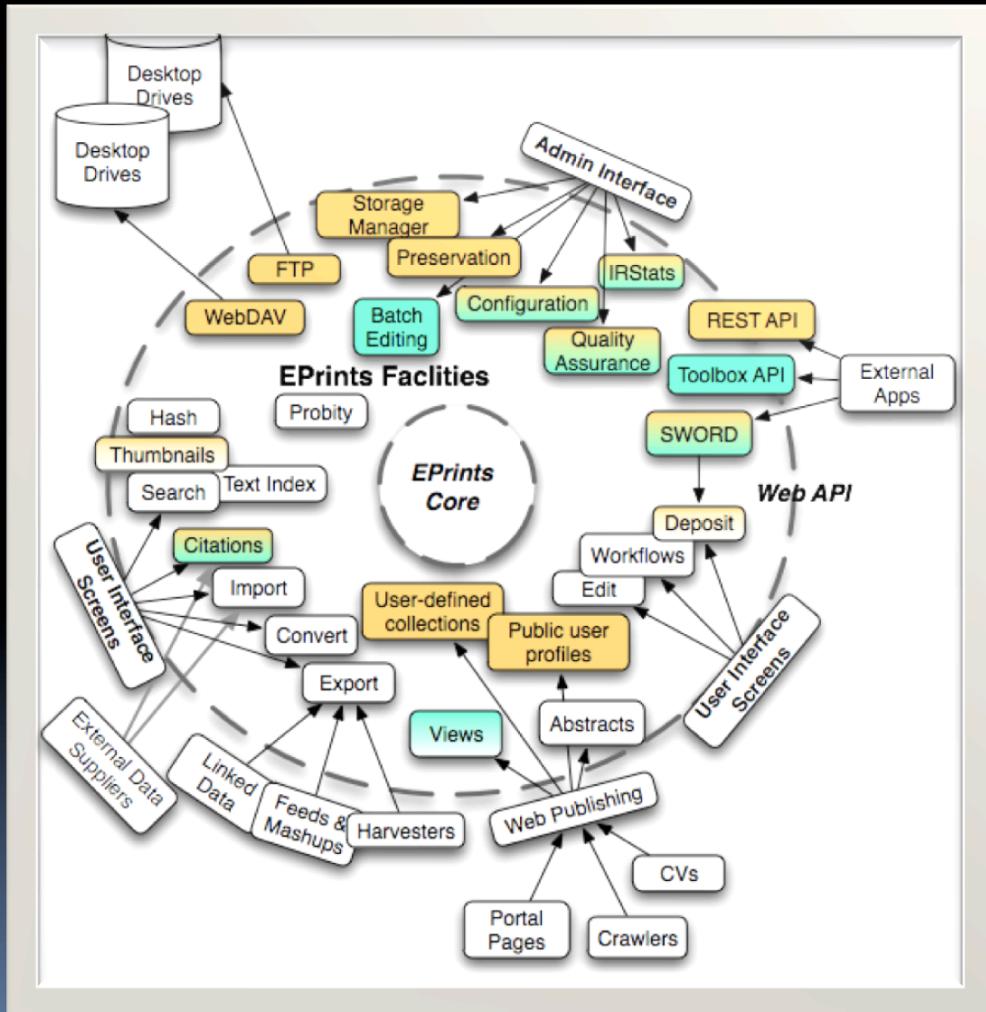
Introduced in 3.1

Introduced in 3.0



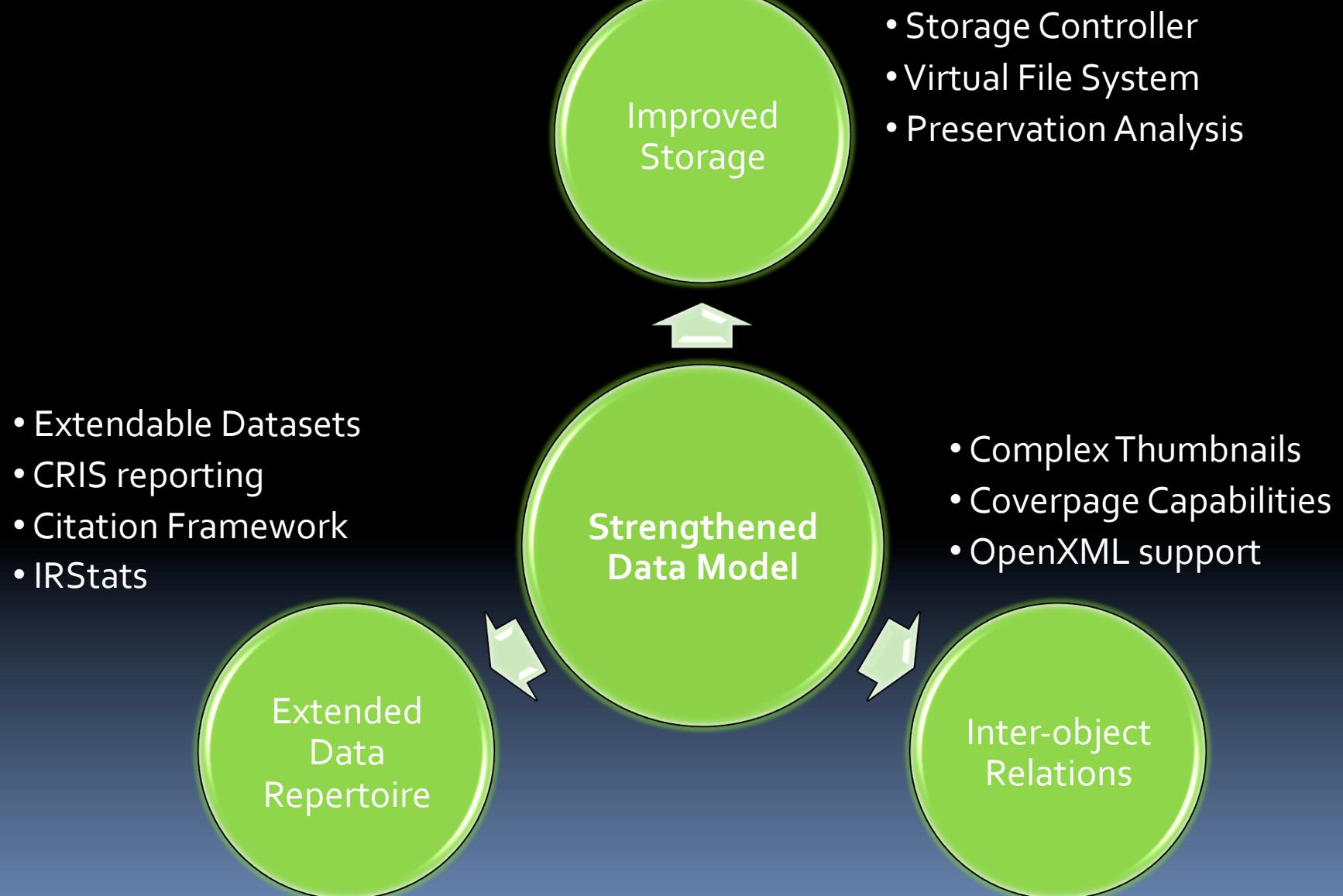
- Datasets
- Data storage
- Utility methods
 - Data types
 - Configuration control
 - Query language
 - Control language *etc*

EPrints Mantle



- External interactions
 - web pages
 - web services
 - other protocols

EPrints 3.2 Improvements



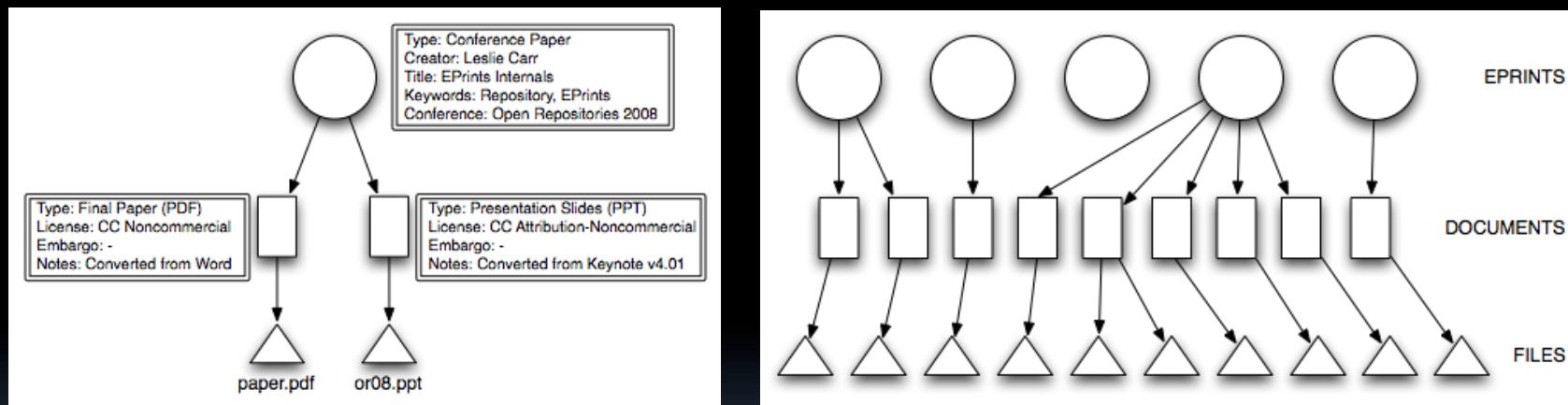
EPrints 3.2 Improvements



See EPrints Wiki for details – http://wiki.eprints.org/w/New_Features_in_EPrints_3.2

Files are now data objects

- Previously, eprints & documents were data objects. Files were just items in a file system.



- Now files are data objects too, and file objects are abstractions of data streams.



Improved
Storage

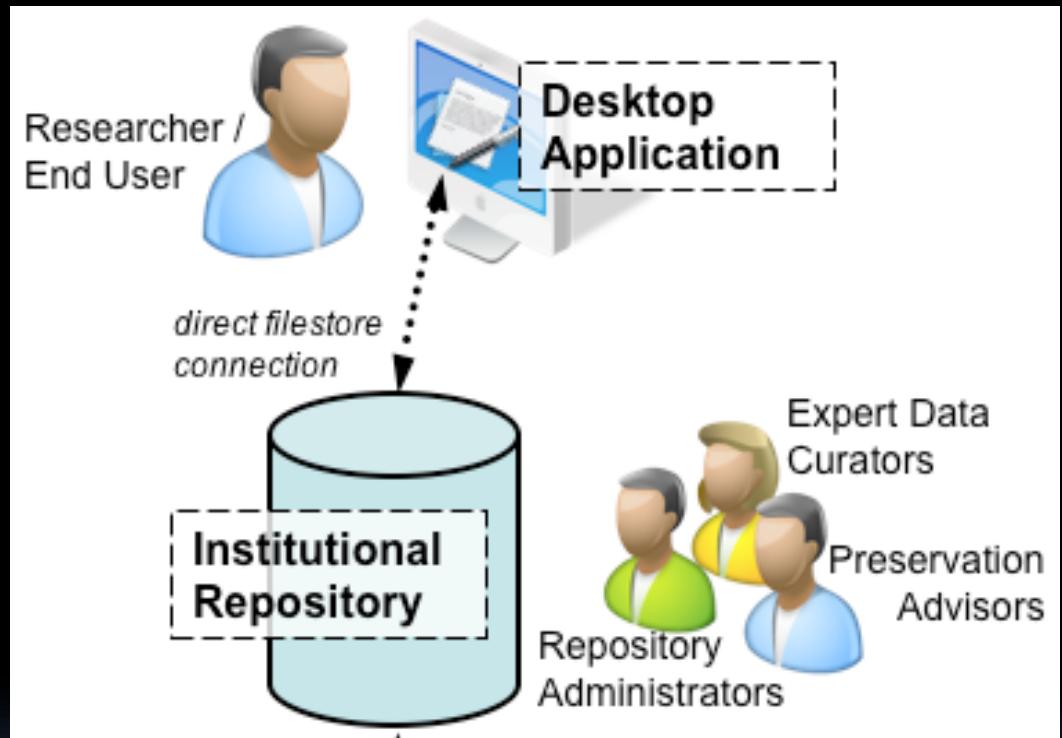
Storage Controller

- Files can be stored on multiple storage services
 - Local disk, SAN, NAS, Honeycomb, Cloud
- Various policies can be enacted through an EPControl XML files.
- Services can be monitored and files swapped between services.

Improved Storage

Virtual File System

- A VFS module invents a view of the repository's files and structures as a hierarchical file system.



- *WebDAV and FTP servers are implemented to allow client desktops to mount the repository as a pseudo-filesystem.*

Complex Thumbnails

- Thumbnails were hidden files
- Now they are documents



- Now thumbnails can be different formats
 - png, jpg, flv...
- Thumbnails can be uploaded manually

Thumbnail Addressing

- Main document
 - <http://eprints.com/123/4>
- Thumbnail
 - <http://eprints.com/123/4.haspreviewThumbnailVersion>
- Chaining Relationships
 - <http://eprints.com/123/4.haspreviewThumbnailVersion.ispreviewThumbnailVersionOf>

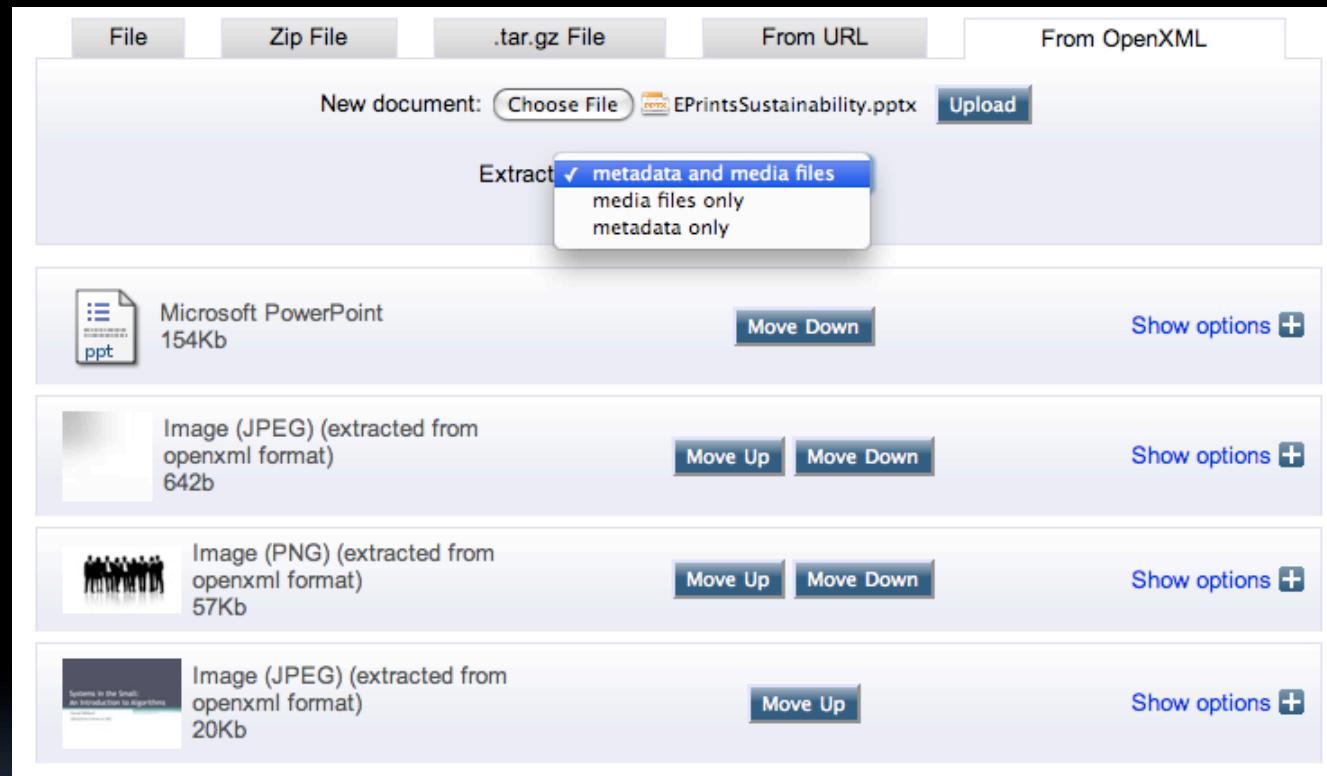
Object IDs, REST & Linked Data

- All data objects have their own data set and id
 - <http://eprints.com/id/eprint/123>
 - <http://eprints.com/id/user/123>
 - <http://eprints.com/id/document/123>
- Type these into a browser URL and EPrints will redirect to the appropriate display object
 - e.g. the eprint that contains document id 123
- Or in other contexts to RDF or other

Object IDs & REST

- EPrints objects are exposed as data for reading and writing through a Web interface
- <http://devel.eprints.org/cgi/rest/eprint/4.xml>
 - The XML output of an eprint
- <http://devel.eprints.org/cgi/rest/eprint/4/creators/2/name/family.txt>
 - text of the surname of the second author of eprint 4

Microsoft Office Support



The screenshot shows the EPrints interface for uploading and extracting files. At the top, there are tabs: File, Zip File, .tar.gz File, From URL, and From OpenXML. The From OpenXML tab is selected. Below the tabs, there is a 'New document:' section with a 'Choose File' button, a file input field containing 'EPrintsSustainability.pptx', and an 'Upload' button. A dropdown menu is open over the 'Extract' button, showing three options: 'metadata and media files' (which is checked), 'media files only', and 'metadata only'. Below this, the uploaded file is listed as 'Microsoft PowerPoint 154Kb'. To the right of the file are 'Move Down' and 'Show options' buttons. Below this are three extracted images: 'Image (JPEG) (extracted from openxml format) 642b', 'Image (PNG) (extracted from openxml format) 57Kb', and 'Image (JPEG) (extracted from openxml format) 20Kb'. Each image has 'Move Up' and 'Move Down' buttons to its left, and a 'Show options' button to its right.

- A new document upload tab allows PPTX/DOCX files to be burst open
 - Media files stored as related documents
 - Metadata added to the eprint
- Also happens with SWORD plugin

Demos

Coming in 3.3+

- Final CERIF extensions
- EPrints Package Manager / App Store
 - Launch at Repository Fringe, Sep 2010
- Scheduler / Calendar interface to event planning
- Multistage Editorial Control
- Manual raising of Quality Issues

Conclusions

- EPrints as a competent and mature part of the researcher's information environment
- Integrating with their desktop knowledge creation activities (making and editing documents)
- Syndicating rich summaries to Web portals, Web 2 and social networking environments
- Participating in linked data / semantic web

Discussion & SWOT Analysis

Strengths

Weaknesses

Opportunities

Threats