Decades of Achievement

<table>
<thead>
<tr>
<th>Internet (packets)</th>
<th>Web (documents)</th>
<th>Repository (preservation)</th>
</tr>
</thead>
</table>

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 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”
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
Open Scientific Data

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
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 Core

- 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

Improved Storage
- Storage Controller
- Virtual File System
- Preservation Analysis

Strengthened Data Model
- Extendable Datasets
- CRIS reporting
- Citation Framework
- IRStats

Extended Data Repertoire
- Complex Thumbnails
- Coverpage Capabilities
- OpenXML support

Inter-object Relations
EPrints 3.2 Improvements

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.
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.
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

- Chaining Relationships
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

- 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
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

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opportunities</td>
<td>Threats</td>
</tr>
</tbody>
</table>