Changing Platforms
Open Repositories 2011

Imma Subirats (FAO of the UN)
Richard M. Davis (ULCC)
Changing Platforms

Parallel case studies of repository platform migration projects

E-LIS: E-prints in Library and Information Science
  ● Large international subject repository (12,000 full texts)
  ● Managed and administered by volunteer effort
  ● Supported and hosted by CILEA
  ● Migrated from EPrints to DSpace in 2010

SAS-Space: School of Advanced Study (University of London)
  ● Small institutional repository (1000-2000 full text articles)
  ● Managed and administered by academic staff
  ● Supported and hosted by ULCC
  ● Migrated from DSpace to EPrints in 2010
E-LIS: EPrints in Library and Information Science
E-LIS on EPrints 2003 - 2010
E-LIS: Migration

Why?
E-LIS: Migration

Technical specialists willing and able to voluntarily host and support the repository were only available from the DSpace community
E-LIS: Migration

How?
E-LIS: Migration

Recreate and reimplement in DSpace....

1. Repository setup
   - Customised layout
   - Structure
   - Communities
   - Collections
   - Workflows
   - Indexes
   - Search options
E-LIS Migration

Co-ordinate...

2. Migration of...
   ● Contents
   ● Metadata
   ● Full-text items (over 11,000)
   ● Statistical data (nearly 13 million records)
   ● Users accounts

3. Redirection of URLs
   ● No persistent item ID in EPrints
   ● New item IDs (Handles) in DSpace
E-LIS Migration: Outcomes

- Content migrated successfully
- User accounts migrated successfully (but different login procedure necessary)
- Continuity of service through upgrade period

- Access to full-text download statistics
- End-user usability issues (e.g. full record view)
- Editor usability issues (e.g. edit-item)
- Workflow less flexible than EPrints
- Fewer administrator functions (e.g. search of all work areas)
- Some EPrints features still unimplemented
E-LIS: Now!
Changing Platforms

Parallel case studies of repository platform migration projects

E-LIS: E-prints in Library and Information Science
● Large international subject repository (11,000 full texts)
● Managed and administered by volunteer effort
● Supported and hosted by CILEA
● Migrated from EPrints to DSpace in 2010

SAS-Space: School of Advanced Study (University of London)
● Small institutional repository (1000-2000 full text articles)
● Managed and administered by academic staff
● Supported and hosted by ULCC
● Migrated from DSpace to EPrints in 2010
SAS-Space: Migration

Why?
SAS-Space: Migration

- Repository needed refreshing: identified need for aesthetic and functional enhancements and improved integration (e.g. LDAP, institutional Web sites)
- Repository manager and users' reservations about DSpace GUI design issues still present in 1.6 - extensive (expensive) customisation work necessary
- Available expertise and experience in ULCC was with EPrints (already hosting or supporting 10 other EPrints repositories)
- EPrints 3.1 instances available to demonstrate features for users and administrators
- Availability of value-add plugins for EPrints: IRStats, SNEEP, MePrints
SAS-Space: Migration

How?
SAS-Space: Migration

- DSpace command line export tool generated packages containing metadata and digital objects
- XSLT and scripts to transform DSpace DC into EP3 XML
- EPrints import tool to load metadata and import data objects
- Scripted and manual validation of import
- DSpace Subjects (uncontrolled keywords) cleaned up and mapped to new EPrints subject tree
- DSpace Communities/Sub-Communities mapped to EPrints Divisions (organisational structure)
- DSpace Collections implemented using additional "subject" field
- Set up LDAP authentication for user accounts
- New visual design commissioned from external Web design team
SAS-Space: Migration

The Handle problem

- DSpace handles Handle 'out of the box'
- EPrints does not have built in Handle support

But do we even need Handle?

We operate the SAS.AC.UK domain. Is it an insurance policy we don't need?

TBL: Cool URIs don't change
http://www.w3.org/Provider/Style/URI
SAS-Space: Migration

Our Handle solution

EPrints import option "enable-import-ids" ensures item ID is the same as in DSpace

$ ./eprints3/bin/import sasspace2dev --enable-import-ids --enable-file-import --force archive XML $item_id/eprints-metadata.xml

Apache redirect - trivial regex in apachevhost.cfg:

RewriteRule ^/dspace/handle/10065/([0-9]+)$ /$1 [R=permanent,L]

Old ones still current, but we don't coin any new ones!
Welcome to SAS-Space
Search results collected from some of the world’s best resources

SAS-Space is an online library for humanities research outputs, providing a permanent archive for scholars and researchers.

Search
ADVANCED HELP

Latest additions
1) Buried Treasure: some lesser-known items in the Institute of Advanced Legal Studies
2) Gross taxation at source: limitations of the international tax framework
3) Duke of Westminster to Vodafone2 - a journey from literal interpretation to liberal rewriting? Conforming construction of UK law in the context of EU direct tax
4) Impact of International Accounting Standards in implementation of national securities regulation
5) Judicial Review: An essential tool for curbing the excesses and abuse of executive action in Sierra Leone
6) Judicial Activism: Usurpation of Parliament’s and Executive's legislative functions, or a Quest for Justice and Social Transformation
Repository platforms: Decisions, decisions

Motivations for choosing repository platform?

- Features
  - Metadata formats
  - User interface
  - Import/export tools
  - Interoperability
  - Searching
  - Reporting
  - Web 2.0 features

- Cost
  - Licencing
  - TCO
Decisions, decisions

JISC Repository Support Project (www.rsp.ac.uk)

Repository software survey, November 2010

Product Comparison Table

<table>
<thead>
<tr>
<th>Feature</th>
<th>CONTENTdm</th>
<th>Digital Commons</th>
<th>DigiTool</th>
<th>DSpace</th>
<th>EPrints</th>
<th>EQUELLA Repository</th>
<th>Fedora</th>
<th>Islandora/Fedora</th>
<th>intraLibrary</th>
<th>Open Repository</th>
<th>Zentity</th>
</tr>
</thead>
<tbody>
<tr>
<td>License Cost:</td>
<td>Commercial</td>
<td>Commercial</td>
<td>Commercial</td>
<td>Free</td>
<td>Free</td>
<td>Commercial</td>
<td>Free</td>
<td>Free</td>
<td>Free</td>
<td>Commercial</td>
<td>Free</td>
</tr>
<tr>
<td>Product Type:</td>
<td>Software</td>
<td>Hosted service</td>
<td>Software</td>
<td>Software</td>
<td>Software</td>
<td>Software</td>
<td>Software</td>
<td>Software</td>
<td>Hosted service</td>
<td>Software</td>
<td></td>
</tr>
</tbody>
</table>

Support

<table>
<thead>
<tr>
<th>Free (community) Support:</th>
<th>-</th>
<th>-</th>
<th>-</th>
<th>Free - community</th>
<th>Free - community</th>
<th>-</th>
<th>Free - community</th>
<th>-</th>
<th>Free - community</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free (direct) Support:</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Free - direct</td>
<td>Free - direct</td>
</tr>
<tr>
<td>Not-for-profit Paid Support:</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Commercial Paid Support:</td>
<td>Commercial</td>
<td>Commercial</td>
<td>Commercial</td>
<td>Commercial</td>
<td>Commercial</td>
<td>-</td>
<td>-</td>
<td>Commercial</td>
<td>Commercial</td>
<td>-</td>
</tr>
<tr>
<td>Update Cost (minor revisions):</td>
<td>With support</td>
<td>With license</td>
<td>With support</td>
<td>Free</td>
<td>Free</td>
<td>With license</td>
<td>Free</td>
<td>Free</td>
<td>Commercial</td>
<td>With license</td>
</tr>
<tr>
<td>Update Cost (major revisions):</td>
<td>With support</td>
<td>With license</td>
<td>With support</td>
<td>Free</td>
<td>Free</td>
<td>With license</td>
<td>Free</td>
<td>Free</td>
<td>Commercial</td>
<td>With license</td>
</tr>
</tbody>
</table>
Repository platforms: Decisions, decisions

Motivations for choosing repository platform?

- Community
  - Available skills and expertise
  - Trusted relationships
  - Sustainability
Changing Platforms: Some Conclusions

EPrints and DSpace have some differences in 'added value' repository features, but both "share the idea of repository (sustainability, accessibility, responsibility)" (Les Carr)

Common approach to essentials makes migration of content (objects and metadata) relatively straightforward

The right solution in any instance is the one that is the best fit to all criteria - including available technical skills and resources.

Avoid platform-related silos: continue to build on common history and heritage of all OS OA platforms to work together towards common goals
Keep Stuff Safe!

(but accessible too)
Credits

Thanks to

- Andrea Bollini (CILEA)
- Dr Peter Webster (Institute of Historical Research, UL)
fin

*