Somaya Langley
University of Cambridge
Polonsky Digital Preservation Project

• Cambridge University Library and Oxford Bodleian Libraries two-year collaborative research project, funded by the Polonsky Foundation

• Three Fellows in each institution: Outreach & Training, Policy & Planning, Technical

http://www.dpoc.ac.uk
Digital Preservation System - Demands

- Integration (with other Library systems)
- Standards-based
- Error handling (automated, prioritised)
- Reporting
- Provenance Notes (including handling provenance, structural and technical metadata that travels with the object into the digital preservation system)
Martin Iordanidis

hbz
Rosetta in a large scale academic environment

hbz’s project aims to support North-Rhine Westfalian universities, with
• different players
• heterogenous materials
• two main goals:
  long-term-availability – re-use being the main objective
• long-term-preservation integrated with all the partner’s usage scenarios
Academic partner requirements

• keep research data within a trusted environment
• cope with big data (if needed)
• open system approach for enhancements
• synergies using a consortial model
Our approach

• hbz operates central Rosetta infrastructure
• **academic data centers may run their own instance**
• part of the regional infrastructure as needed
• stakeholder’s interest...
  – data ownership
  – big data scalability
• ...and **responsibilities**
  – local libraries and data centers provide knowledge transfer and support researchers
pilot phase

• setting up three partner’s use cases with ExL support
• use cases from
  – University Of Cologne Library
  – University Of Cologne Data Center (RRZK)
  – RWTH Aachen
• hbz: sandbox and production system
• RRZK: production only
• knowledge transfer to content producers
• Rosetta training phase
Thank You!

Paul Kirschner, Martin Iordanidis & Andres Quast
teamrosetta@hbz-nrw.de
State Library of Queensland
Problem 1 – 260 Disk Image Archive

Michael Bryce Archive

- signage for the Brisbane 1982 Commonwealth Games
- the logo for the Government of Queensland
- fashion brands Fendi and Valentino
Problem 1 – 260 Disk Image Archive

Michael Bryce Archive (cont’d)

- Doha Stadium
- the Eurostar train
- San Pellegrino
- International Cricket Council
- Harrods, London
Problem 1 – 260 Disk Image Archive

Michael Bryce Archive is a major digital archive

Problem is that it is all on 260 disks
Problem 1 – 260 Disk Image Archive
Problem 1 – 260 Disk Image Archive
Problem 1 – 260 Disk Image Archive
Problem 1 – 260 Disk Image Archive
Problem 1 – 260 Disk Image Archive

The document "fbh_con_mp1c.indd" uses fonts that are currently unavailable on your computer. Closing the dialog box will substitute the missing fonts with the default font until the original fonts become available.

Typekit couldn’t find any matching fonts.

<table>
<thead>
<tr>
<th>Missing Fonts</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Helvetica Neue 77 Bold Condensed:</td>
<td>Substituted with default font.</td>
</tr>
<tr>
<td>Helvetica Neue 35 Thin:</td>
<td>Substituted with default font.</td>
</tr>
<tr>
<td>Helvetica Neue 55 Roman:</td>
<td>Substituted with default font.</td>
</tr>
<tr>
<td>Helvetica Neue 75 Bold:</td>
<td>Substituted with default font.</td>
</tr>
</tbody>
</table>
DEFINITION OF FILM FROM OUR REGION
Problem 1 – 260 Disk Image Archive
Problem 1 – 260 Disk Image Archive

- Like to be able to store all of the environmental information into the IE but not as descriptive metadata

- Eventually like to connect the IE to an ‘thing’ which can create the required environment using the PREMIS environmental information
Problem 2 – Alma Integration

• SLQ requires Alma to be the source-of-truth for all descriptive metadata

• So IE level descriptive metadata is replicating well (except collections info)
Problem 2 – Alma Integration

- Would like selective FL descriptive metadata to be replicated into Alma

- FL to be replicated would be based on IE type (e.g. Photo albums with individually described images)
Yvonne Tunnat
ZBW
Migration action without a trace?

Rosetta Preservation Planning

Rosetta Advisory Group
12th June 2017
Sheffield
presenter: Yvonne Tunnat, ZBW Kiel / Germany
The action

PDF -> PDF/A

The outcome

<table>
<thead>
<tr>
<th>Title</th>
<th>Version</th>
<th>Commit Date</th>
<th>Created By</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kliniken in Baden-Württemberg...</td>
<td>2</td>
<td>10/05/2017 08:34:28</td>
<td>friese@...</td>
<td>Events View METS Download</td>
</tr>
<tr>
<td>Kliniken in Baden-Württemberg...</td>
<td>1</td>
<td>14/01/2017 03:01:12</td>
<td>SubAppZBW</td>
<td>Events View METS Download</td>
</tr>
</tbody>
</table>
The wish
get the information from the xml report into the Rosetta metadata

Ideas:
• Same report with IE
• Save as PREMIS event
• Transfer to DNX metadata

But how?
Euwe Ermita
State Library of New South Wales
Digital Preservation team and capabilities
Questions

Digital Preservation team

• Who has a dedicated digital preservation team?
• Centralised or decentralised
• Size of team
• Types of roles (Manager, Developer, TA, Strategy, Sysadmin, Application support?)
• Organic or strategically established

Capabilities

• Who has an active digital preservation training program?
Carl Wilson
Open Preservation Foundation
A Brief Introduction to the OPF

Carl Wilson
Rosetta Advisory Group, Sheffield 2017
About us

● International not-for-profit membership organisation
● Hosted by the British Library, Yorkshire
● 3 staff
● 21 members in Europe and North America
  ○ archives | libraries
  ○ research and technology organisations
  ○ companies providing digital preservation services

shared solutions for effective and efficient digital preservation
Sustaining project results

- Technology
- Knowledge
- Communities
Open source file format identification, validation & characterisation

- Under OPF stewardship since 2015
- Long established digital curation tool developed by JSTORE & Harvard and maintained by Gary McGath
- Format validation for 14 formats
- Used in Rosetta and other repository software
- OPF focussing on code quality and automated testing
- Latest release: V 1.16, April 2017

http://jhove.openpreservation.org/
https://github.com/openpreserve/jhove
Open Source, industry supported PDF/A validation

- Funded by the EU PREFORMA project (PCP)
  http://www.preforma-project.eu/
- Validation of all PDF/A flavours 1b, 1a, 2b, 2a, 2u, 3b, 3a, 3u
- Feature extraction (characterisation) of PDFs
- Custom policy enforcement beyond ISO 19500
- Basic metadata repair, e.g. PDF/A flags in XMP
- License: GPL v3+ and MPL v2+
- Latest release: 1.6 June 2017

http://verapdf.org/  https://github.com/veraPDF  @_veraPDF
Technology

- Software quality
  - Software maturity model (based on Apache, ISO25000)
  - Github (code repository & bug reports)
  - Travis (continuous builds)
  - Jenkins (automated tests & deployment)
  - Codacy & CodeCov (code quality monitor)
- R&D projects
- Tool stewardship
- Affiliate member contributions
  - new Fido release coming soon (Artefactual)
- Code review (BitCurator, TIMBUS, EARK)
Join - help - support the OPF

- All our software is open source
  - file bug reports
  - provide fixes
  - help write documentation
  - add to corpora

- Our knowledge and information resources are open to all
  - write blog posts
  - join activities
  - use resources

- Support the foundation
  - charter member
  - affiliate member
  - software supporter (currently JHOVE - future veraPDF, other?)
Thank you!

carl@openpreservation.org

http://openpreservation.org/

@openpreserve

Illustrations from digitalbevaring.dk
Ross Spencer
Archives New Zealand
Archives NZ - Doing More with More Preservation Tools

Lightning (and Thunder) Talk ;)

We have *all* the tools available to us...

Exiftool

Apache Tika

VeraPDF

SaxonXML

Custom Scripts and Software...

Just not inside Rosetta...
But Ex Libris are helping, SDK, Developer's Blog, GitHub etc. Setting an example…

SLUB, LDS Church, {Your name here}

But what is the evolution of a plugin?

What does the knowledge economy look like?

- Finding limits and capabilities of preservation plugins
- **AKB** - MD requirements for Namespaces - Incorporating multiple tools doing the same thing
- Tools and guides to help reduce the cost of development
- Open to more than the ‘developers’ - or is it all too hard/expensive to begin?
- Feedback loops - testing
- When new customers arrive for Rosetta, they get more than DROID/Jpylyzer/JHOVE
- The Acorn? - Minimum API? How do we develop guides for external makers (OPF?) to create for us?
Archives NZ wants to be involved more...

Combination of cost, resources, capabilities, make it difficult... So we want to hear those stories too...

Three WIBINI (Wouldn’t It Be Nice If) Plugins

- **Character encoding checker** x-fmt/111 is good, but an encoding is better!

- **Embedded object evaluation**, e.g. PDF embedded in PPT! **Preservation**!

- **Missing object evaluation**, i.e. what should be there?! - Missing links, missing files, e.g. WMV external to a PPT...
FIN.
Teresa Soleau
Getty
Sharing is caring
(for our cultural heritage)

PROBLEM:

9 TB of born-digital files languishing on a server uncharacterized or preserved

We use METS deposit for digitized content which requires server access, training and knowledge of XML. Needed a more lightweight way for digital archivists to deposit content.

We couldn't use the out of the box ZIP to CSV function because we needed to include checksums we already had and local provenance notes (related to file renaming).

No in-house software development to write custom tool.
Sharing is caring (for our cultural heritage)

SOLUTION:

Worked with Ross Spencer at ANZ around 2014 to test and edit his code for CSV deposits.

RESULT:

Digital archivist can do all the steps herself.

After loading born-digital content from institutional archives into Rosetta for the past 6 months we have over 530,000 files in 26 IEs preserved in the system and growing steadily.

Ross was able to see it “in the wild” which helped him to streamline it and make it more robust - helping all users both internal and external.