Open Preservation

Community approaches to ensuring long-term access to digital collections
The OPF team

Community Manager
- Events (face-to-face/virtual)
- Training (staff development)
- Comms (web/email/social)

Executive Director
- Membership (engagement/value)
- Open preservation advocacy
- Operational management

Technical Lead
- Infrastructure (host/test)
- Software stewardship (roadmap/maturity/packaging)
- Data corpora
Software Stewardship
Knowledge Exchange
Services
Advocacy and Alliances
Overview

- Preservation Scenarios: scale and perspective
- State of the Art: software and practices
- Repository Integration: tools and user experience
- The Role of Community: shared knowledge + shared effort == sustainable solutions
- Opportunities for Collaboration
Preservation Scenarios: small scale

- Characterisation and validation
  - Profiling heterogeneous collections (e.g. deposited archives)
  - Analysing disk image contents (e.g. legacy media)
  - Dealing with email formats (e.g. institutional records)
- Detecting duplicates (e.g. collection digitisation projects)
- Automated metadata extraction
Preservation Scenarios: small scale
Preservation Scenarios: med - large

- Automation (e.g. quality control of format migrations)
- Scalability (rapidly growing collections)
- Diverse sources (e.g. validating 3rd party submissions)
- Complexity (variety of formats and files, e.g. digital rights management, digital art, or non-conformance to specs)
Preservation Scenarios: the future?
State of the Art: common tools

- Identify
- Characterise
- Validate
- Analyse
State of the Art: SCAPE tools

Characterisation and Quality Assurance:

- pagelyzer - web page rendering
- matchbox - image analysis (duplicates, cropping)
- jpylyzer - JPEG 2000 validator
- xcorrsound - audio waveform analysis
State of the Art: SCAPE systems

- PLATO (preservation planning)
- SCOUT (preservation watch)
- Platform (parallel processing)
- C3PO (content profiling)
State of the Art: SCAPE ecosystem

- Automated Watch
- Automated Planning
- Component Management
- Digital Object Repository
- Execution Platform
Repository Integration

- RODA (Fedora 3)
- Fedora 4 (reference SCAPE integration)
- Rosetta (Ex Libris)
- Hydra (digital preservation working group)
Repository Integration

● Eprints (preservation plugins)
● Preservica (Tessella)
● microservices / workflows
Repository Integration: make it easy

Improving usability of digital preservation tools

- Discoverability (easy to find)
- Packaging (easy to install)
- Documentation (easy to use)
Community: Open Knowledge

Shared Knowledge

- reports
- case studies
- blogs
- wikis
Community: Open Knowledge

Community Owned digital Preservation Tool Registry (COPTR)

COPTR describes tools useful for preserving digital information for the long term. COPTR is also an initiative to collate the knowledge of the digital preservation community on preservation tools in one place. Instead of organisations competing against each other with their own registries, COPTR is bringing them together. In doing so, its objective is to provide the best resource for practitioners on digital preservation tools.

Browse the COPTR Registry
- View all tools
- View tools by functional category
- View tools by the type of content they act upon

How to create a new Tool Entry
1. Check out these Guidelines for contributing to COPTR.
2. Search for the tool you want. Check the full name and the acronym!
3. If you find it, consider adding more detail to the existing entry. If you don’t find it, follow the link to create a new entry, which will automatically create an outline entry for your tool based on the Tool Page Template.

COPTR Partners

<table>
<thead>
<tr>
<th>Partner</th>
<th>Logo</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCC</td>
<td><img src="image" alt="DCC Logo" /></td>
</tr>
<tr>
<td>DCE</td>
<td><img src="image" alt="DCE Logo" /></td>
</tr>
<tr>
<td>NDSA</td>
<td><img src="image" alt="NDSA Logo" /></td>
</tr>
<tr>
<td>Open Planets Foundation</td>
<td><img src="image" alt="Open Planets Foundation Logo" /></td>
</tr>
<tr>
<td>POWRR</td>
<td><img src="image" alt="POWRR Logo" /></td>
</tr>
</tbody>
</table>

The Digital Curation Centre (DCC)
The Digital Curation Exchange (DCE)
National Digital Stewardship Alliance (NDSA)
The Open Planets Foundation (OPF)
Preserving digital Objects With Restricted Resources (POWRR)
Community: Open Knowledge

Digital Preservation

Business Case Toolkit

● making the case
● benefits, risks, costs…
● templates, FAQ, case studies
Community: Open Knowledge

Hack Events:

- Traditionally face-to-face (developers and practitioners)
- Start with a tell and show / speed dating format
- Followed by development on specific issues
Community: Open Software

- Shared Requirements
  - Wiki Pages
  - Issue Logs
  - Testability

- Shared Experiences
  - Wiki pages
  - Blogs
Community: Open Software

- Shared Code
  - Open Source
  - Visibility
  - Project quality
Community: Open Software

- Shared Testing == Robust Tools
  - Continuous integration
  - Community corpora
  - Functional testing
  - Regression testing
  - Published results
  - Towards continuous deployment
Community: Partnerships
Community: Partnerships

Memory Institutions (charter members)

- Sustainable software
- Knowledge exchange
- Development efficiency
- Agile/risk tolerance

Experience Requirements Testing Data

New communities (tool adoption/testing)

Global benefits
- sustainability
- data broker
- multiple communities
- TRUST/CONFIDENCE

Open Planets Foundation

Research and Development (affiliates, commercial)

Visibility/Adoption Sustainability Requirements Testing

Software Theory

Sustainable software Reputation

Service Providers (repository systems)
Opportunities for Collaboration

- Digital preservation can seem technically complex, what are the needs for integrating available tools and services with widely adopted repository systems?
- Solutions can appear achievable only at scale (large institutions with dedicated teams), what are the requirements for deployable solutions at all scales of operation?
- The repository and preservation communities can seem isolated from each other, how can we ensure better collaboration on shared challenges?
My Details

- Carl Wilson
- Technical Lead
- Open Planets Foundation
- Email: carl@openplanetsfoundation.org
- Skype: carl.f.wilson
- GitHub: carlwilson
- Twitter: @openplanets
- Google+: carl@openplanetsfoundation.org