Digital Preservation Community

OPF/SCAPE Executive Seminar
02 April 2014, The Hague

Ed Fay, Executive Director
ed@openplanetsfoundation.org | @digitalfay
Memory Institutions (Charter Members)

Knowledge exchange
Robust tools
Development efficiency
Agile/risk tolerance

Experience
Requirements
Testing
Data

Tools and theory

Research and Development (affiliates, commercial)

Visibility/Adoption
Sustainability
Requirements
Testing

Robust tools
Reputation

New communities (tool adoption/testing)

Service Providers (Ex Libris, Tessella...)

Open Planets Foundation
Memory Institutions (Charter Members)

Knowledge exchange
Robust tools
Development efficiency
Agile/risk tolerance

Experience
Requirements
Testing
Data

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

Tools and theory

Research and Development (affiliates, commercial)

Visibility/Adoption
Sustainability
Requirements
Testing

New communities (tool adoption/testing)

Robust tools
Reputation

Service Providers (Ex Libris, Tessella…)

Open Planets Foundation
based on H. Balk - Pennington de Jong (2013)
Centres of Competence and Virtual Centres of Excellence
Open-source → Commercial
Open-source

Workflow

Implementation

Operations

Institution

Open-source

Commercial

User groups

Community

Best practice

Standards

Peer-audit

Research
Mission/Values

Open
- knowledge sharing, open source licensed technology

Inclusive
- memory institutions (national library/archive)
- diversification? (museums, galleries, commercial)
- smaller organisations? (.ac/.edu, local/regional archives)

Trustworthy
- confidence building (tools/approaches/policy)
- member profiles (sharing expertise visibly)
- leadership role (technology and practice)
Strategic themes

Knowledge exchange
- Now: events, webinars, case studies (blogs/wiki)
- Future: tech reports, best practices (formalising knowledge)

Tool stewardship
- Now: infrastructure, testing (and reporting), corpora
- Future: “stewardship” - efficient development, confidence-building

Services

Advocacy/alliances
- Research and development - growing software portfolio
- Digital Preservation Coalition - joint events, tech watch
- EU - consultation responses, project matchmaking
Knowledge Exchange

- PDF/A (forthcoming...)
- JHOVE2
- Preserving your preservation tools
- SCAPE tools and services (several, some forthcoming...)
- Hadoop/scalability
- Securing funding for digital preservation
- Preservation planning
- File format identification and characterisation
- Digital library development and practice at LSE (that was me!)
- Application infrastructure for action services (workflow automation)
- Forensics tools and methods (disk images and BitCurator)
- Emulation-as-a-service (CD-ROM ingest and scientific process)
- Content profiling (C3PO)
- Software development and collaborative best practices
- Command-line digital preservation tools
- Preservation health check (using PREMIS)
- Understand the value of preservation data
- Practical tools and approaches to preservation systems
- Database archiving (SIARD, MIXED, RODA)

events, webinars
Knowledge Exchange
Knowledge Exchange

Welcome to the OPF Knowledge Base Wiki

The OPF wiki is here to gather knowledge relating to key issues in digital preservation and organize these to allow practitioners, developers and information specialists to better preserve digital heritage. This wiki welcomes contributions from every member of the community and provides a hosting space for many global projects. From practitioners, we welcome your use case scenarios and problems and hope to link these with solution providers. For developers, we not only hope to join you to practitioners in the community through online and face-to-face events, but also welcome your valuable solutions with advice on software development, project hosting and distribution. Gathering together communities and key knowledge allows information managers to make effective decisions for preserving our digital heritage.

Contribute your requirements and solutions
Share your challenges and solutions with the community, and view the resulting record of digital preservation practice.

Collaborate
Get involved with these collaborative digital preservation initiatives, including Q&A on Stack, the Atlas of Digital Damages and the OPF Format Corpus.

Find preservation tools
Find tools for solving your digital preservation challenges in the OPF Tool Registry.

Understand File Format Risks
File format risks you should be concerned about, and how to identify them in your content.

Develop Software with the OPF
All you need to know about developing software with the OPF, with guidance on making your software effective, sustainable, maintainable and reusable.

OPF Projects
Projects hosted by OPF, including: Aqua, SCAPE, SPRUCE and Preservation Health Check.

OPF Events and Webinars
The latest OPF events and hackathons.

Join the conversation
Blog, comment and exchange ideas on the main Open Planets Foundation website.

Sign up for the OPF wiki

Datasets, Issues and Solutions

OPF Format Corpus

Atlas of Digital Damages

Community infrastructure
Knowledge Exchange

Welcome to the OPF Knowledge Base Wiki

The OPF wiki is here to gather knowledge relating to key issues in digital preservation and organize those to allow practitioners, developers and information specialists to better preserve digital heritage. This wiki welcomes contributions from everyone member of the community and provides a meeting space for many global projects. From practitioners, we welcome your use case stories: only hope to join you to practice valuable solutions with advice communities and key knowledge heritage.

Blogs

Developing an Audio QA workflow using Hadoop: Part II

First things first. The Github repository with the Audio QA workflows is here: https://github.com/statsbiblioteket/scape-audio-qa. And version 1 is working. Version is really all wrong here. I should call it Workflow 1, which is this one:

Submitted by Bolette Ammitzb... on 3 February 2014 - 11:33am

Read more Bolette Ammitzbhal Julk's blog

Why can't we have digital preservation tools that just work?

One of my first blogs here covered an evaluation of a number of format identification tools. One of the more surprising results of that work was that out of the five tools that were tested, no less than four of them (FITS, DROID, Plaso and jHDFView) failed to even run when executed with their associated launcher script. In many cases the Windows launcher scripts (batch files) only worked when executed from the installation folder. Apart from making things unnecessarily difficult for the user, this also completely flies in the face of all existing conventions on command-line interface design. Around the time of this work (summer 2011) I was in contact with the developers of all the evaluated tools, and until last week I thought those issues were a thing of the past. Well, was I wrong!

Submitted by Johan van der Knijff on 31 January 2014 - 12:58pm

Read more Johan van der Knijff's blog

SCAPE survey on preservation monitoring. Participate now!

Anyone willing to preserve digital content must be aware of events that might constitute a relevant risk. In SCAPE we are developing tools that will allow you to detect risks before they cause any irreversible damage.

Help us understand the preservation events, threats and opportunities, you find more relevant and the ways you would like us to detect them.

Participate in our survey and help us develop tools that would help you to automatically detect problems in your own content, and events that might put it at risk.
Knowledge Exchange

Welcome to the OPF Knowledge Base Wiki

The OPF wiki is here to gather knowledge relating to key issues in digital preservation and organize these to allow practitioners, developers and information specialists to better preserve digital heritage. This wiki welcomes contributions from every member of the community and provides a hosted space for many global projects. From practitioners, we welcome your use case scenarios, only hope to join you to practice valuable solutions with advice communities and key knowledge heritage.

Blogs

Community Owned digital Preservation Tool Registry (COPTR)

First things first. The GitHub repository is the best place to go if you want to contribute to COPTR. 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

Why can't we

One of my first blogs has the more surprising result than four of them (YTS), associated launches script when executed from the user, this also complies Interface design. Around evaluated tools, and until

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

- DCC
- dce
- NDSA
- Open Planets Foundation
- POWRR

SCAPE survey

Anyone willing to preserve risk in SCAPe we are an irreversible damage. Help us understand the ways you would like to participate in our survey and help us to understand how you maintain secure access to your own data, and events that might put it at risk.
Tool stewardship

EU research
- SCAPE, PLANETS
- tools, prototypes, best practice

Affiliate/Commercial members
- r&d institutions
- e.g. emulation, forensics

Charter members
- framework for collaborative development/testing
  $\rightarrow$ efficiency, consistency
- future: shared priorities/roadmaps? e.g. characterisation
LSE Digital Library
LSE Digital Library

Senior Management
• Deputy Director

Digital Library Team
• Digital Library Manager
• Software Developer x 2
• Systems specialist
• User Experience Architect

Academic/Archive Services
• Collection Librarians
• Repository Manager
• Digital Archivist

Collection Services
• Collections Manager
• Metadata/Discovery Mgr
• Metadata Tech Officer
• Senior Library Assistant
• Library Assistant

Central IT Department
• work with various posts
LSE Digital Library

Senior Management
- Deputy Director

Digital Library Team
- Digital Library Manager
- Software Developer x 2
- Systems specialist
- User Experience Architect

Academic/Archive Services
- Collection Librarians
- Repository Manager
- Digital Archivist

Collection Services
- Collections Manager
- Metadata/Discovery Mgr
- Metadata Tech Officer
- Senior Library Assistant
- Library Assistant

Central IT Department
- work with various posts

New posts
New skills
LSE Digital Library

Senior Management
- Deputy Director

Digital Library Team
- Digital Library Manager
- Software Developer x 2
- Systems specialist
- User Experience Architect

Academic/Archive Services
- Collection Librarians
- Repository Manager
- Digital Archivist

Collection Services
- Collections Manager
- Metadata/Discovery Mgr
- Metadata Tech Officer
- Senior Library Assistant
- Library Assistant

Central IT Department
- work with various posts
  (networking, infrastructure, info security)

Strategy

Management

Collection development

Ingest

Discovery

Preservation (policy, planning)

Applications

Virtual environments

User research, prototyping

Metadata profiles, crosswalks
Workforce development

Digital preservation
- DPTP (high-level, institutional perspective)
- OPF/DPC events (current and emerging practice)
- Repository collaborations (Hydra project)

User Experience
- agency contracts (3 x projects, full lifecycle)
- new, permanent post (hired via agency)
- shifting internal/external balance (user research, information architecture, prototype/test shifted in-house; visual design and front-end dev remained external)
Business Case Toolkit

• Step-by-step
prepare yourself!, audit readiness, assess where you are and what you need, think about your audience, work out your objectives, benefits + alignment, sort your documentation, validate and refine, deliver with impact (politics...), share!

• Template
executive summary, strategic vision (where is this going? how long will it take?), background, context, and status (environment, state of the collection(s), institutional readiness, stakeholders, preservation risks), business activity (what are you proposing? what are the options?), value for money/ROI (benefits, costs, implementation risks)

• FAQ
when is the right time? why are we doing this? what resources are we focusing on? who is affected? how do I make the case?

• Case studies
Bishopsgate Institute, Institute of Education, Northumberland Estates, Lovebytes, University of St. Andrews
Ed Fay, Executive Director
ed@openplanetsfoundation.org | @digitalfay