Digital Preservation Interoperability through Preservation Actions Registries

Matthew Addis, Justin Simpson, Jon Tilbury, Jack O’Sullivan, Paul Stokes, Carl Wilson

JISC
Arkivum, Arefactual, Preservica, Open Preservation Foundation

http://parcore.org/presentations/
Agenda

• Background & Motivation
• Project outcomes
• Proof of concept implementation
• Next Steps
Users want the best advice, wherever it comes from:
- Identification, property extraction, validation, migration, rendering, tools

Multiple parallel initiatives research and advise on best practice:
- Products such as Preservica & Archivematica
- Practitioners
- Academics
- Specialists

but they don’t talk to each other effectively.
Background: Motivation and Objectives

• Want to
  o Improve the quality and ease of use of advice sent to practitioners as soon as it is available
  o Improve research cooperation and reduce repetition

• Expected outcome
  o Provide a mechanism to exchange information between all parties regardless of which system they use

• Exclusions
  o Protocols for prioritising and authorising which advice applies to which user / system / intent
  o One registry to rule them all
Background: Jisc RDSS Project

Development of a multi-vendor shared services platform drove discussions of interoperability of format policies (i.e. “preservation actions”) between preservation systems.
Background: Project Conception

A JISC funded project to initiate the process to deliver benefits to RDSS users

Arkivum, Preservica and Artefactual as RDSS product suppliers

Open Preservation Foundation as respected independent shared DP technology supplier
1) Preservation is not just about file formats, it’s about intellectual entities/objects

2) We need a way to define/describe context - why is this action being taken? what is the business rule?

3) Reinventing the wheel - preservation actions are not portable across systems (e.g. Archivematica, Preservica, others)
Background: PAR Impacts

<table>
<thead>
<tr>
<th>GOAL</th>
<th>ACTOR</th>
<th>IMPACT</th>
<th>DELIVERABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empowered to make informed preservation decisions</td>
<td>Digital Preservation Professional</td>
<td>Information about formats &amp; actions are easy to find &amp; trust</td>
<td>search &amp; access capabilities</td>
</tr>
<tr>
<td>Accelerated Adoption</td>
<td>Funding institutions (HEIs)</td>
<td>Preservation actions are easy to understand and drill into</td>
<td>Data Model that defines actions, tools, formats, properties</td>
</tr>
<tr>
<td>Confidence in preservation actions</td>
<td>Jisc</td>
<td>Preservation actions can be assessed for quality &amp; adoption</td>
<td>Indicators that demonstrate &quot;best practices&quot; in the community</td>
</tr>
<tr>
<td>Decrease cost &amp; risks of ownership</td>
<td>Preservation System Vendor</td>
<td>Easy to switch vendor products or use multiple products</td>
<td>API to publish &amp; consume from registries</td>
</tr>
<tr>
<td>Improve Preservation System Quality</td>
<td></td>
<td>Community best practice incorporated into product</td>
<td>Protocols for incorporating actions into workflow</td>
</tr>
</tbody>
</table>

Governance and change mgmt process to maintain data model & API
Community collaboration features
Versioning & provenance of preservation actions
Registry of preservation actions & info
<table>
<thead>
<tr>
<th>What have we produced and why?</th>
</tr>
</thead>
</table>
| **Conceptual Model** | ● Common framework for everyone to work to  
● Something to argue about and agree upon!  
● Interlingua between preservation systems |
| **Json Schemas** | ● Formal definition of the PAR model  
● Machine readable, used in API payloads  
● Used to test and validate interoperability |
| **API** | ● Common interface for preservation systems  
● Well defined way to exchange information |
| **Executable DP Actions** | ● Cross-platform way to deploy/run tools  
● Unambiguous and vendor independent |
| **Proof of Concept** | ● Prove PAR is possible!  
● Not just a talking shop or paper exercise  
● Reference implementation to share |
PAR Conceptual Model

https://doi.org/10.6084/m9.figshare.6628418
JSON schemas

- Tool
- Action
- Action Type
- Format
- Property
- Business Rule

https://github.com/JiscRDSS/rdss-par/tree/master/schemas
**Preservation Actions**

**Retrieve all preservation actions**

Allow to retrieve the details of all the preservation action

**QUERY PARAMETERS**

- `limit` (string): An integer to limit the number of preservation action returned. Default value is zero, which will no filter the result.
- `offset` (string): An integer to specify the offset of the first element of the list of preservation action returned. Default value is zero, which will no filter the result.
- `modifiedAfter` (string): Filter the preservation action to return only the ones having a localLastModifiedDate value GREATER than the one passed in. Allowed datetime formats are YYYY-MM-DD/THH:mm:ssZ and YYYY-MM-DD. In the second scenario, when the time is not provided, it will be assumed to be 00:00:00 UTC time.
- `modifiedBefore` (string): Filter the preservation action to return only the ones having a localLastModifiedDate value LOWER than the one passed in. Allowed datetime formats are YYYY-MM-DD/THH:mm:ssZ and YYYY-MM-DD. In the second scenario, when the time is not provided, it will be assumed to be 00:00:00 UTC time.

**HEADER PARAMETERS**

- `tool` (string): Filter the preservation actions by the Tool they use, provided its ID. Multiple IDs can be passed in within the same string, linked together using an ampersand symbol as follows: ID_1&ID_2&ID_3. When more than one ID is provided, the endpoint will return all the preservation actions that use one of the IDs in the list. When used together with the preservation_action_type header parameter, the endpoint will return only those preservation actions satisfying both filters' criteria. A list of Tool IDs cannot be combined with another list of Preservation Action Type IDs.

---

https://github.com/JiscRDSS/rdss-par/tree/master/api
**Executable Tool Definitions**

- **Machine readable spec for running a tool**
  - Tool command line
  - Parameters and flags
  - Inputs and outputs
  - Pre and post processing

```json
[job mediaInfo2.cwl] completed success
{
  "width": "1280",
  "bitrate": "748253",
  "height": "720"
}
Final process status is success
```

Fixity check

```json
[job md5check2.cwl] completed success
{
  "fixity_report": "PASS"
}
Final process status is success
```

https://github.com/JiscRDSS/rdss-par/tree/master/examples/cwl
Registry (In)compatibility

Preservica Registry

- File Format
- Property
- Type
- Migration Pathway
- Tool Purpose
- Tool
- Tool Option
- applies to
- has
- applies to
- executes a
- controlled by

Archivematica FPR

- File Format
- Version
- Group
- Rule
- Purpose
- Command
- Tool
- applies to
- has
- executes a
- is for a particular preservation
- using a
Common Language

Diagram showing relationships between file format, property, preservation action, parameters, tool, and business rule.
API Exposure

PAR Preservation Actions

ID: 1ca929de-315-44ac-8d83-51a8a712726
Description: Extraction of properties for Video files using MediaInfo
Tool: mediaInfo
Version: 18.03
Example Constraints:
commandline 'mediaInfo -Output=EBUCore inputFile'

Inputs:
None

Outputs:
None

Status: CONVERGED
FPR Role: 0d4a0E1-163fa-455b-a6e9-62718a43a699
FPR Command: 1ca929de-315-44ac-8d83-51a8a712726
Working API Examples

Demo API Servers:

http://parcore.dev.archivematica.org:62080 (user: test, apikey: test)
http://52.209.71.78/Registry/par (user: test, password: test)
Next steps

- Real use cases demonstrating real benefits
- Consortium of funders
- OPF coordination
- More members to provide API endpoints
- More sync tools to exchange information between systems
- Ability to act on information exchanged
Resources

• Project pages
  o http://www.parcore.org/

• Github repo
  o https://github.com/JiscRDSS/rdss-par/

• iPRES paper
  o https://doi.org/10.6084/m9.figshare.6628418

• DPC blog post
  o https://www.dpconline.org/blog/a-new-era-in-collaboration-in-digital-preservation-research

• Project announcement and contacts
Interested?

This could be the first step in a global exchange of best practice between all stakeholders in DP

We want you to make sure this is truly global.

We’ll be at après-iPRES
Thursday 27 September 1-5pm
Kotzen Room, Lefavour Hall at Simmons College

Vote for PAR to find out more!

Contact: info@parcore.org