



# 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

# Background: The problem

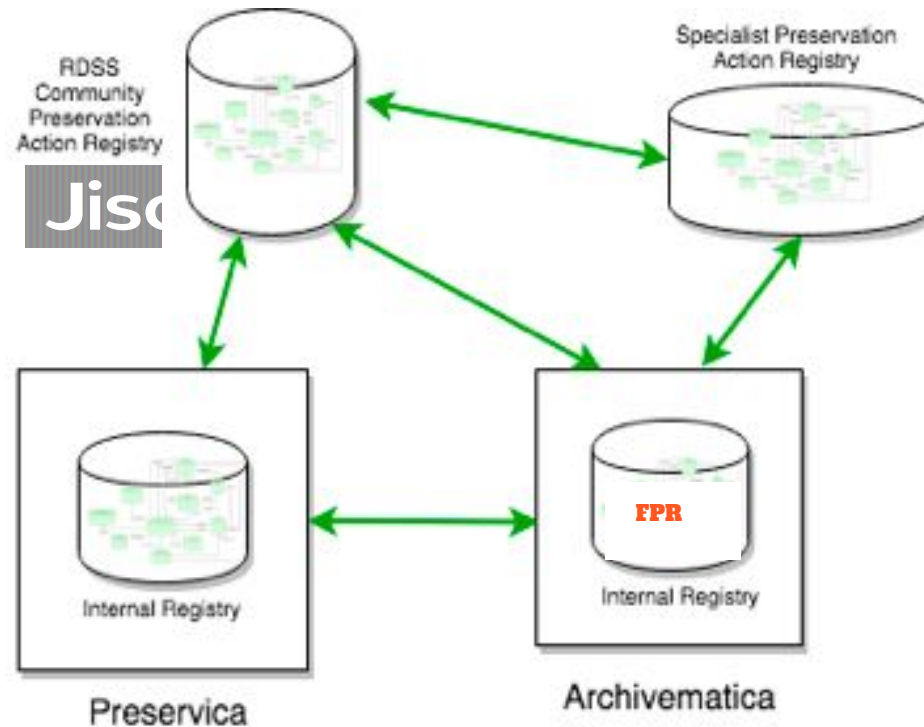
- Users want the best advice, wherever it comes from
  - Identification, property extraction, validation, migration, rendering, tools
- Multiple parallel initiatives research and advise on best practice
  - Products such as Preservica & Archivematica
  - Practitioners
  - Academics
  - Specialists
- but they don't talk to each other effectively

# Background: Motivation and Objectives

- Want to
  - Improve the quality and ease of use of advice sent to practitioners as soon as it is available
  - Improve research cooperation and reduce repetition
- Expected outcome
  - Provide a mechanism to exchange information between all parties regardless of which system they use
- Exclusions
  - Protocols for prioritising and authorising which advice applies to which user / system / intent
  - 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

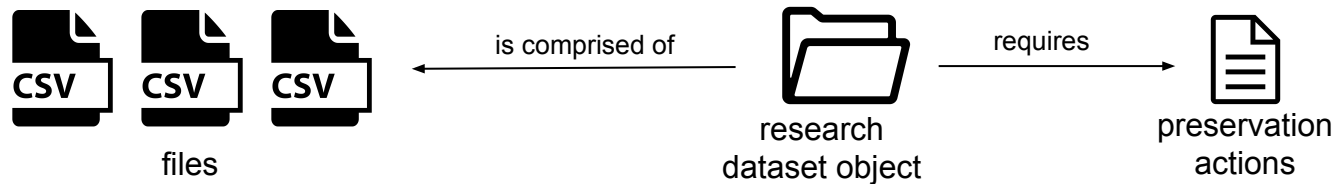
A JISC funded project to initiate the process to deliver benefits to RDSS users

Arkivum, Preservica and Artefactual as RDSS product suppliers

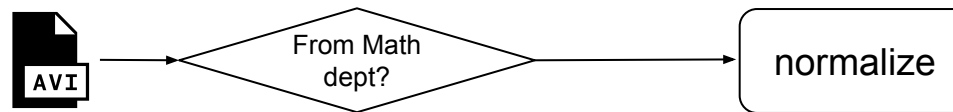
Open Preservation Foundation as respected independent shared DP technology supplier

# Background: Motivations

- 1) Preservation is not just about file formats, it's about intellectual entities/objects

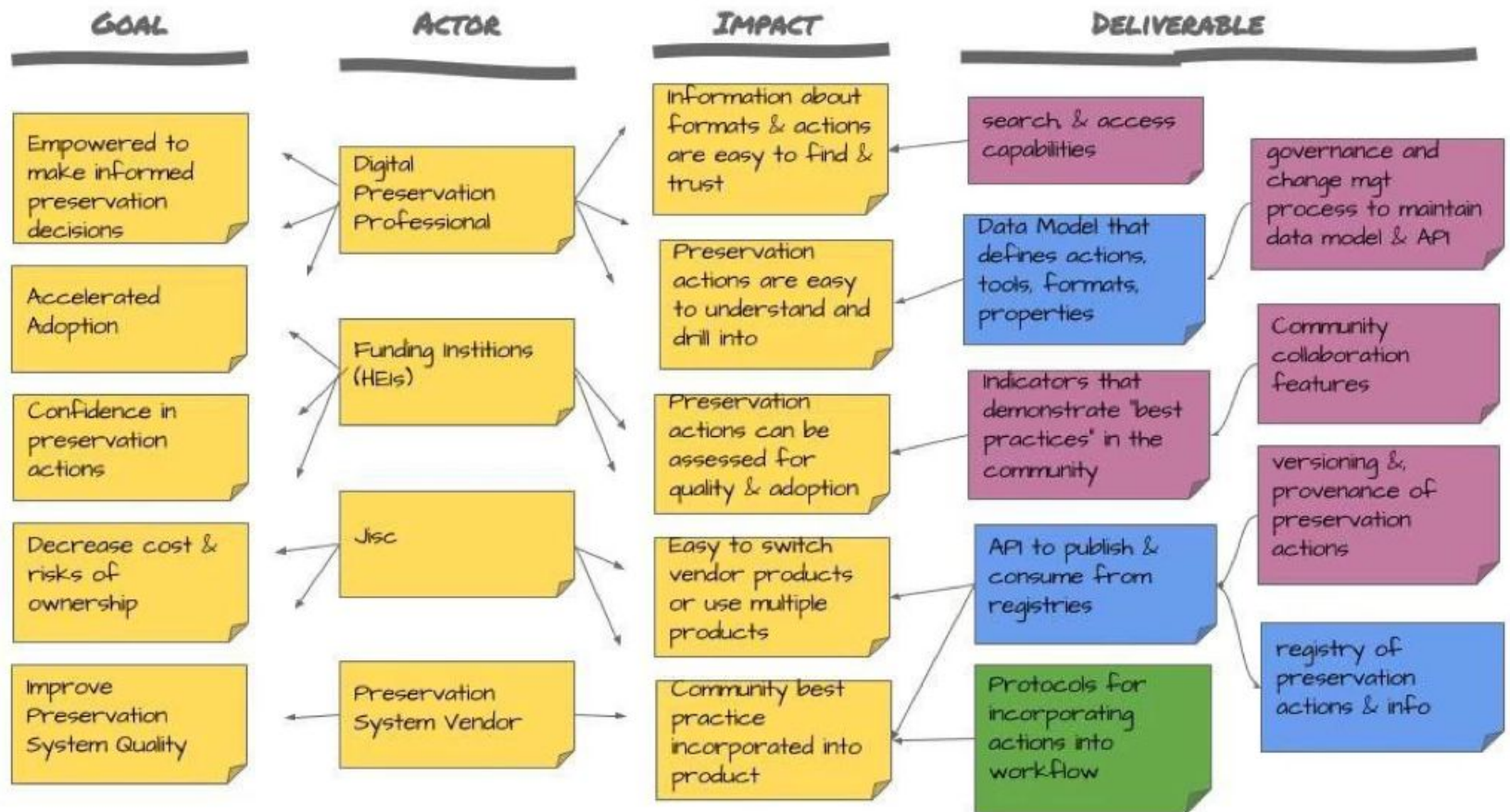


- 2) We need a way to define/describe context - why is this action being taken? what is the business rule?



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

# Background: PAR Impacts

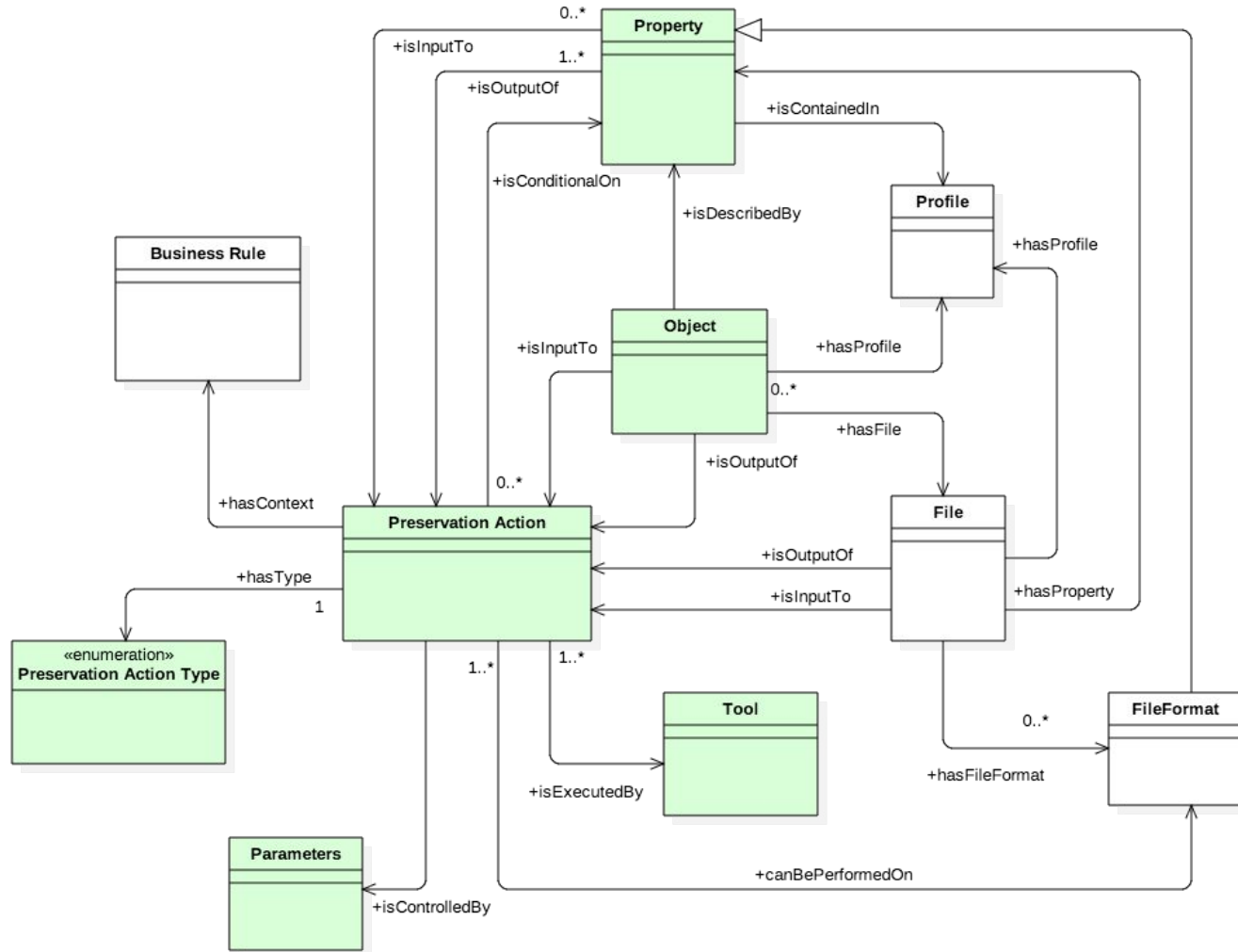




# What have we produced and why?

Conceptual Model	<ul style="list-style-type: none"><li>• Common framework for everyone to work to</li><li>• Something to argue about and agree upon!</li><li>• Interlingua between preservation systems</li></ul>
Json Schemas	<ul style="list-style-type: none"><li>• Formal definition of the PAR model</li><li>• Machine readable, used in API payloads</li><li>• Used to test and validate interoperability</li></ul>
API	<ul style="list-style-type: none"><li>• Common interface for preservation systems</li><li>• Well defined way to exchange information</li></ul>
Executable DP Actions	<ul style="list-style-type: none"><li>• Cross-platform way to deploy/run tools</li><li>• Unambiguous and vendor independent</li></ul>
Proof of Concept	<ul style="list-style-type: none"><li>• Prove PAR is possible!</li><li>• Not just a talking shop or paper exercise</li><li>• Reference implementation to share</li></ul>

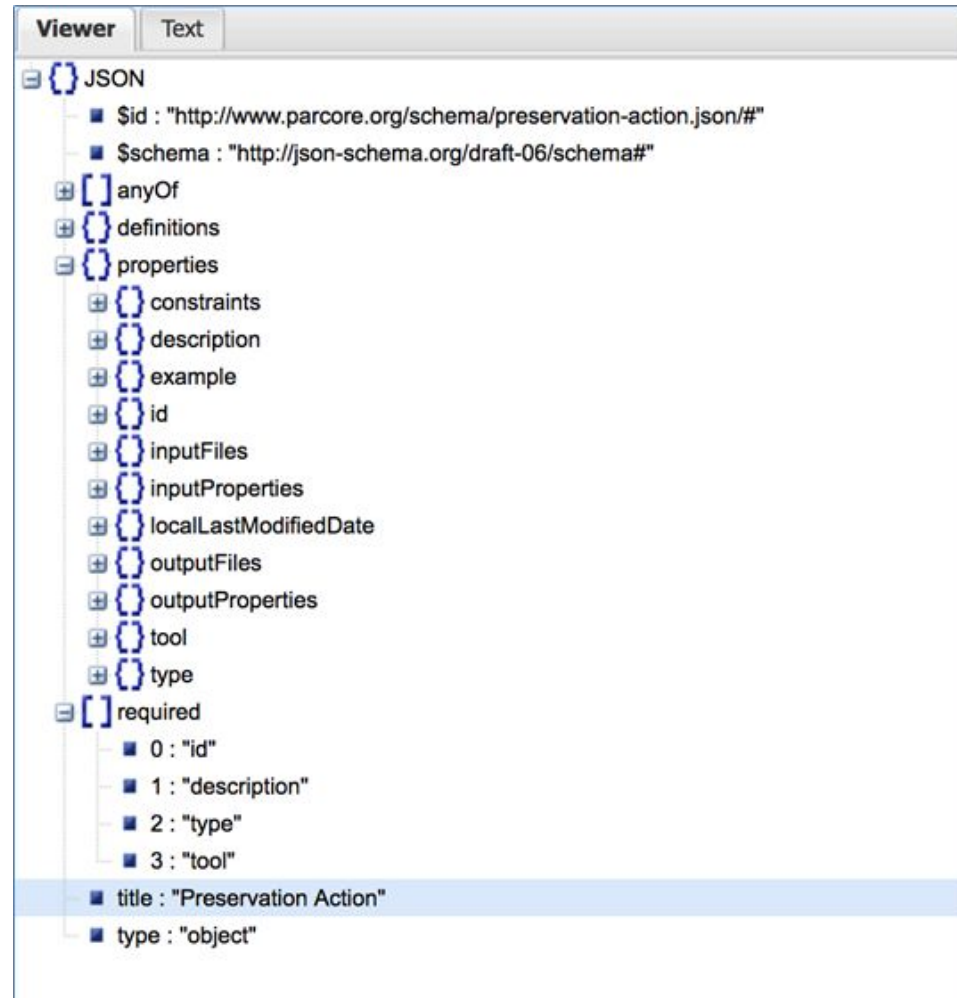
# PAR Conceptual Model



<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>

# APIs

The screenshot displays the RDSS API documentation for the 'Preservation Actions' endpoint. The left sidebar contains a navigation menu with categories: AUTHENTICATION, PRESERVATION ACTION REGISTRY, BUSINESS RULES, FILE FORMATS, PAR PROPERTIES, PRESERVATION ACTION TYPES, and PRESERVATION ACTIONS (selected). The main content area is titled 'Preservation Actions' and 'Retrieve all preservation actions'. It includes a description: 'Allow to retrieve the details of all the preservation action'. Below this, there are sections for 'QUERY PARAMETERS' and 'HEADER PARAMETERS'. The 'QUERY PARAMETERS' section lists: 'limit' (string, integer to limit results), 'offset' (string, integer to specify offset), 'modifiedAfter' (string, filter by localLastModifiedDate), and 'modifiedBefore' (string, filter by localLastModifiedDate). The 'HEADER PARAMETERS' section lists: 'tool' (string, filter by tool ID). On the right, a 'Response samples' section shows a '200' status code and a JSON response sample for 'application/json'.

**Preservation Actions**

**Retrieve all preservation actions**

Allow to retrieve the details of all the preservation action

**QUERY PARAMETERS**

Parameter	Type	Description
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\T\hh:mm:ssTZD 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\T\hh:mm:ssTZD 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**

Parameter	Type	Description
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 follow: 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.

**Response samples**

200

application/json

```
{
  "constraints": [
    + { ... }
  ],
  "description": "string",
  "example": "string",
  "id": {
    "guid": "string",
    "name": "string",
    "namespace": "string"
  },
  "inputFiles": [
    + { ... }
  ],
  "inputProperties": [
    + { ... }
  ],
  "localLastModifiedDate": "2018-07-31T16:03:25Z",
  "outputFiles": [
    + { ... }
  ],
  "outputProperties": [
    + { ... }
  ],
}
```

<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



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

Property extraction

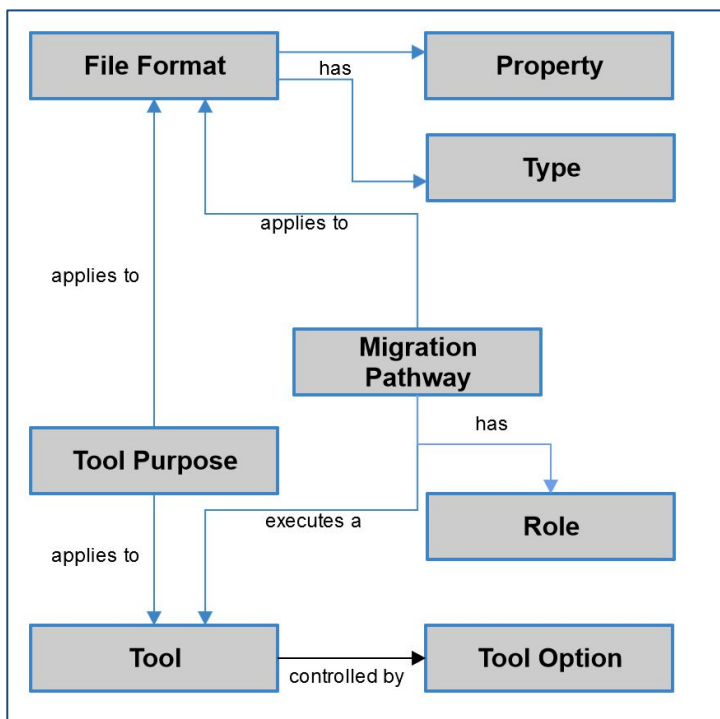
Fixity check

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

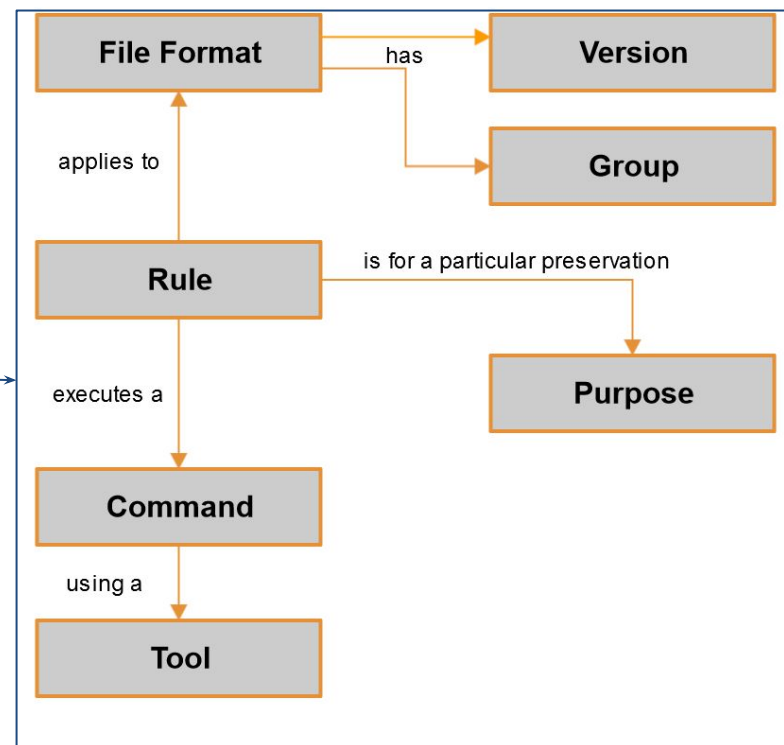
<https://github.com/JiscRDSS/rdss-par/tree/master/examples/cwl>

# Registry (In)compatibility

Preservica Registry

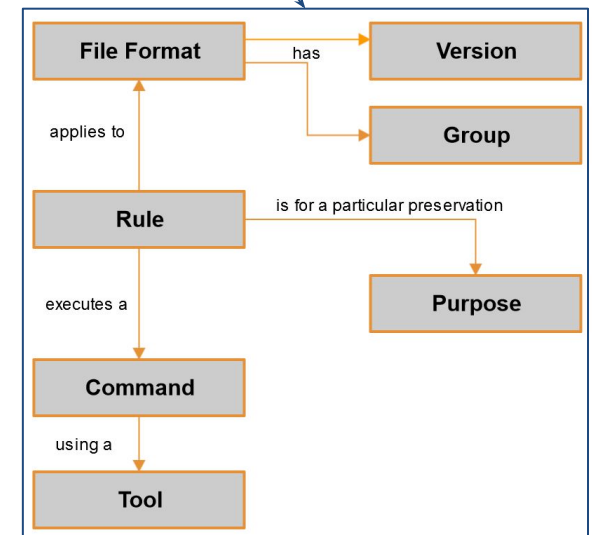
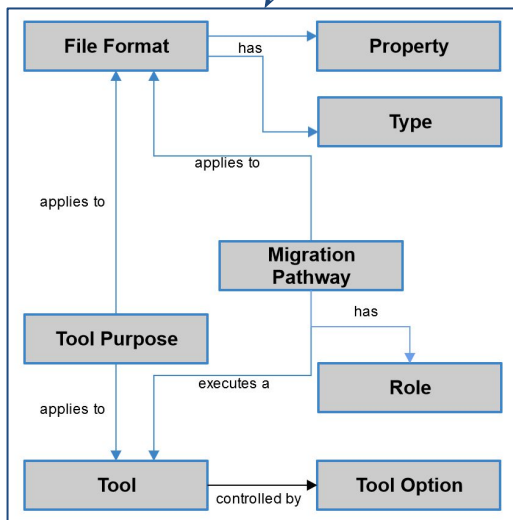
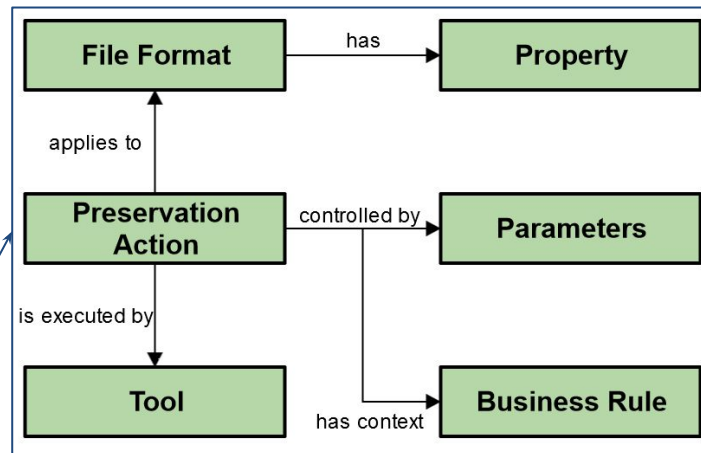


Archivematica FPR



?

# Common Language





# API Exposure

← → ↺ 🏠

**archivematica** Transfer Backlog Appraisal Ingest Archival storage Preservation planning Ac

PAR / Preservation Actions / List

FPR PAR Preservation Actions

## PAR Preservation Actions

<b>Id</b>	1caa0cde-e345-44ac-8d83-51afaa7427b6
<b>Description</b>	Extraction of properties for Video files using MediaInfo
<b>Type</b>	metadata extraction
<b>Tool</b>	mediainfo
<b>Version</b>	18.03
<b>Example</b>	commandline 'mediainfo --Output=EBUCore inputFile'
<b>Constraints</b>	<pre>[   {     "inputItemName": "inputFile",     "allowedFormats": [       {         "localLastModifiedData": "2018-06-14T15:00:00.00Z",</pre>
<b>Inputs</b>	None
<b>Outputs</b>	None

**Status** CONVERTED

**FPR Rule** 5e54cf31-637a-4f58-a859-03f184ab5f49

**FPR Command** 1caa0cde-e345-44ac-8d83-51afaa7427b6

← → ↺ 🏠

JSON Raw Data Headers

Save Copy Collapse All Expand All

▼ preservationActions:

▼ 0:

▼ constraints:

▼ 0:

allowedFormats: []

allowedPropertiesAllOf: []

allowedPropertiesAnyOf: []

▼ description: "Extraction of properties for Video files using MediaInfo"

▼ example: "commandline 'cwltool mediaInfo2.cwl mediaInfo2.yml'"

▼ id:

guid: "13a27a64-0671-525e-8d18-d01b62de2849"

name: "mediainfo2"

namespace: "http://par.preservica.com"

▼ inputFiles:

▼ 0:

description: "File that will have metadata extracted from"

▼ file:

filepath: ""

name: "inputfile"

▼ outputProperties:

▼ 0:

description: "Height of the video frame"

name: "height"

▼ parProperty:

▼ id:

guid: "9c00f7d7-99f0-5efe-bc65-ae02dbc5a05d"

name: "height"

namespace: "https://www.ebu.ch/metadata/ontologies/ebucore"

type: "integer"

class: "size"

units: "pixels"

▼ 1:

description: "Bitrate of the video file"

name: "bitrate"

▼ parProperty:



# Working API Examples

The screenshot shows the Insomnia API client interface. The title bar reads "Insomnia (PAR Core) – PAR Preservation Action object to update an fpr.FPRule". The main interface is divided into several sections:

- Left Sidebar:** Contains a list of API endpoints with their methods:
  - PUT PAR Preservation Action object to update an fpr.FPRule
  - POST PAR format object to create an fpr.FormatVersion
  - POST PAR tool object to create an fpr.FPTool
  - PUT x PAR Object tool to update an fpr.FPTool
  - GET All File Formats
  - GET Specific file format
- Top Bar:** Shows the method "PUT", the URL "base\_url /api/beta/par/preservation\_actions/111bec2e-", and a "Send" button. Status indicators show "201 CREATED", "TIME 163 ms", and "SIZE 138 B".
- Tab Bar:** Includes tabs for "JSON", "Auth", "Query", "Header", and "Docs".
- URL Preview:** Displays the full URL: "http://parcore.dev.archivematica.org:62080/api/beta/par/preservation\_actions/111bec2e-e387-4ab9-8e95-86ce7af6adb c?username=test&api\_key=test".
- Preview Panel:** Shows the JSON response: 

```
1 {
2   "message": "Preservation action successfully updated.",
3   "uri": "/api/beta/par/preservation_actions/111bec2e-e387-4ab9-8e95-86ce7af6adb c"
4 }
```

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
  - <http://www.parcore.org/>
- Github repo
  - <https://github.com/JiscRDSS/rdss-par/>
- iPRES paper
  - <https://doi.org/10.6084/m9.figshare.6628418>
- DPC blog post
  - <https://www.dpconline.org/blog/a-new-era-in-collaboration-in-digital-preservation-research>
- Project announcement and contacts
  - <http://openpreservation.org/news/arkivum-artefactual-the-open-preservation-foundation-and-preservica-collaborate-on-new-jisc-initiative-for-sharing-preservation-action-best-practice/>

# 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](mailto:info@parcore.org)