



Collaborative Preservation with PAR

Arkivum, Artefactual, Preservica, Open Preservation Foundation

PAR overview and features of Release 0.3 - 29 April 2020

Agenda

Item	Presenter
Introduction and Overview	OPF
PAR community stewardship	OPF
PAR use cases	OPF
An overview and demo of the PAR 0.3 release	Jon , Matthew, Jack
Plans for our online workshop	Justin
Q&A	

Introduction & Overview

Today's webinar is the first of three of a series of PAR presentations:

1. Today's webinar

Will focus on what's new in release 0.3 and how you can get involved in our 2nd presentation which will be an interactive workshop based event

2. Workshop

We expect to run the workshop sometime around June. It will be your opportunity to help shape the next and future releases of PAR by helping us to define development features. In addition the workshop will include a review of the PAR data model as well as group work aimed at describing preservation policies using PAR

3. PAR roadmap webinar

The third and final presentation in this series will highlight:

- details of PAR release 0.4 - influenced by the workshop
- details of how you can get actively involved with PAR including development opportunities
- the PAR roadmap

Introduction & Overview

Presenting today are members of the PAR core team including:

- Jon Tilbury and Jack O'Sullivan, Preservica
- Matthew Addis, Arkivum
- Justin Simpson, Artefactual Systems
- Becky McGuinness, Carl Wilson and Martin Speller, OPF

Agenda

Item	Presenter
Introduction and Overview	OPF
PAR community stewardship	OPF
PAR use cases	OPF
An overview and demo of the PAR 0.3 release	Jon , Matthew, Jack
Plans for our online workshop	Justin
Q&A	

PAR Community Stewardship

- Under the stewardship of Jisc, PAR has moved forward from a standing start to its current position that includes a feature-rich API, an endpoint browser with a prototype API implementation spanning Preservica, Archivemata and Arkivum
- With release 0.3 now on a firm footing the ongoing stewardship of PAR is to be transferred to the OPF
- This change is aimed at:
 - providing a more independent technical oversight and overall task coordination of PAR activities, and
 - To expand the community knowledge and involvement with PAR via:
 - regular webinars and workshops
 - encouraging involvement via hack events, and
 - in some cases, as the providers of best practice trusted sources

PAR Community Stewardship

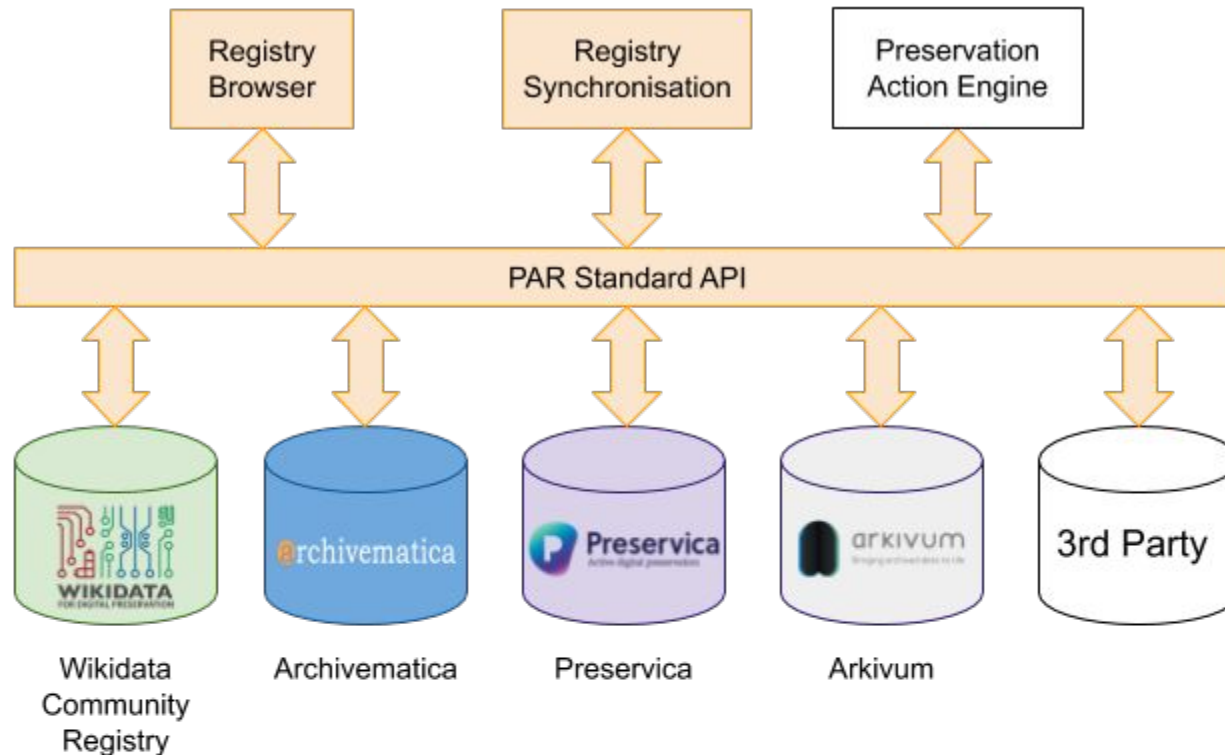
Other OPF stewardship tasks include:

- The creation of a new Github repo that will replace the existing rdss_par repo (with openpreserve/parcore). The new repo will act as a central repository for:
 - all source material
 - PARCORE.org website
 - PAR API documentation (generated from the Swagger source)
- Other focus areas include:
 - creation a new 'open' Slack channel (parcore)
 - community involvement
- Finally, a special word of thanks to Jisc for their support in getting PAR off the ground and for their ongoing help and encouragement

Agenda

Item	Presenter
Introduction and Overview	OPF
PAR community stewardship	OPF
PAR overview and use cases	OPF
An overview and demo of the PAR 0.3 release	Jon , Matthew, Jack
Plans for our online workshop	Justin
Q&A	

PAR Overview



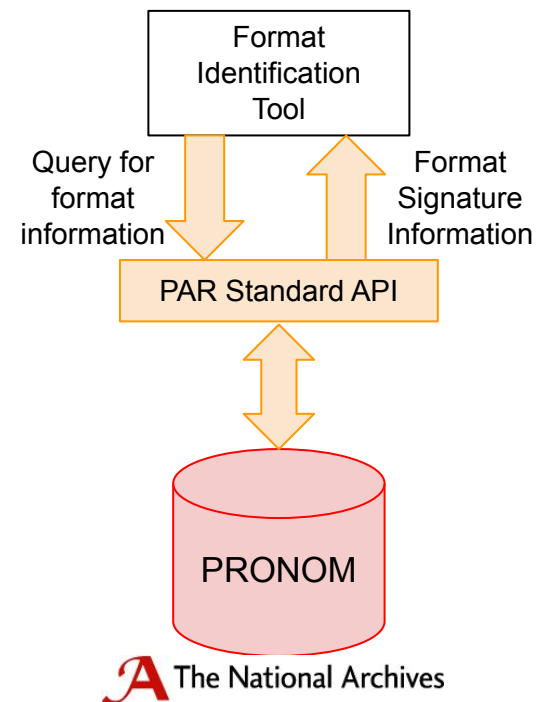
By way of introduction the ultimate purpose of PAR is to:

- facilitate sharing preservation actions between disparate DP systems
- allow well-formed preservation workflows, such as a set of compounded preservation actions, to be imported from an external trusted source for use within your local environment

PAR Use Cases: Updating format signatures

DROID and FIDO are file format identification tools that rely on format signature data that is downloaded from the UK National Archives PRONOM service. While this data is of high quality there can be long gaps between signature file releases meaning it can take a significant time for new format signatures to make it into DROID or FIDO. A PAR based solution would involve:

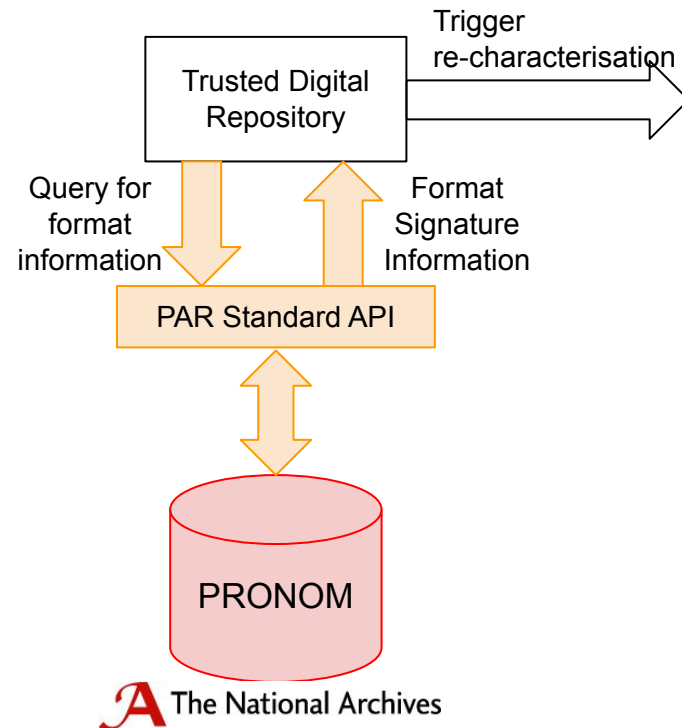
- Implementing a PAR endpoint for PRONOM allowing PAR clients to query and obtain format data.
- Developing a new signature update tool for DROID and FIDO to use the PAR API to query for and download new format signatures.
- Users could choose alternative PAR registries as sources of format information.



PAR Use Cases: Re-Characterise a Collection

One of the many problems facing memory institutions is that improvements to digital preservation tools and practises can mean having to revisit previously ingested material. A simple example is the update of format signatures, making it possible to determine the format of previously unidentified files. PAR compliant tools can be used to detect this situation and automatically re-characterise collections:

- A repository system uses the PAR API to query a signature for new format information.
- When new signatures are available they're automatically downloaded and the identification tool is updated.
- The updated tool is automatically run against unidentified files to try to ascertain their format.



PAR Use Cases: Learning from good practice

A mature PAR ecosystem enables anyone to explore tools and actions used by other organisations facing similar problems. Digital preservation novices could learn from national institutions, preservation specialists, etc.

- Preservation tools and actions can be chosen from trusted sources of good practice
- These can be deployed and tested on PAR action engines as well as production systems.

Agenda

Item	Presenter
Introduction and Overview	OPF
PAR community stewardship	OPF
PAR use cases	OPF
An overview and demo of the PAR 0.3 release	Jon, Matthew, Jack
Plans for our online workshop	Justin
Q&A	

PAR Components

- File Formats
- Tools
- Preservation Actions
- Business Rules

Swagger API Documentation

```
x-content-type-options: nosniff
x-frame-options: DENY, DENY
x-ua-compatible: IE=Edge
x-xss-protection: 1; mode=block
```

Responses

Code	Description
------	-------------

200	
-----	--

Success

Example Value | Model

application/json;charset=UTF-8

```
{
  "fileFormats": [ {
    "description": "This is an outline record only, and requires further details, research or authentication to provide information that will enable users to further understand the format and to assess digital preservation risks associated with it if appropriate.",
    "externalSignatures": [ {
      "signature": "f4a",
      "signatureType": "File Extension"
    }, {
      "signature": "m4v",
      "signatureType": "File Extension"
    }, {
      "signature": "m4a",
      "signatureType": "File Extension"
    }, {
      "signature": "f4v",
      "signatureType": "File Extension"
    }, {
      "signature": "mp4",
      "signatureType": "File Extension"
    }
  ],
  "id": {
    "guid": "1d6dc249-b131-5492-bab2-60d98fa73e02",
    "name": "fmt/199",
    "namespace": "http://www.nationalarchives.gov.uk"
  },
  "internalSignatures": [ {
    "byteSequences": [ {
      "byteSequenceId": "bs/357",
```

POST /file-formats Create FileFormat

Standardised GUI

Archivematica using the same GUI

[Home](#)

Preservation Action Details

[Setup](#)

Property	Value
Name	Extracting 7Zip compatible format.
ID	b8864e2c-87a0-4aec-a8cf-a747a073be28
Namespace	https://archivematica.org
Description	Extracting 7Zip compatible format.
Last Modified	2013-11-15 01:18:36
Type	extraction

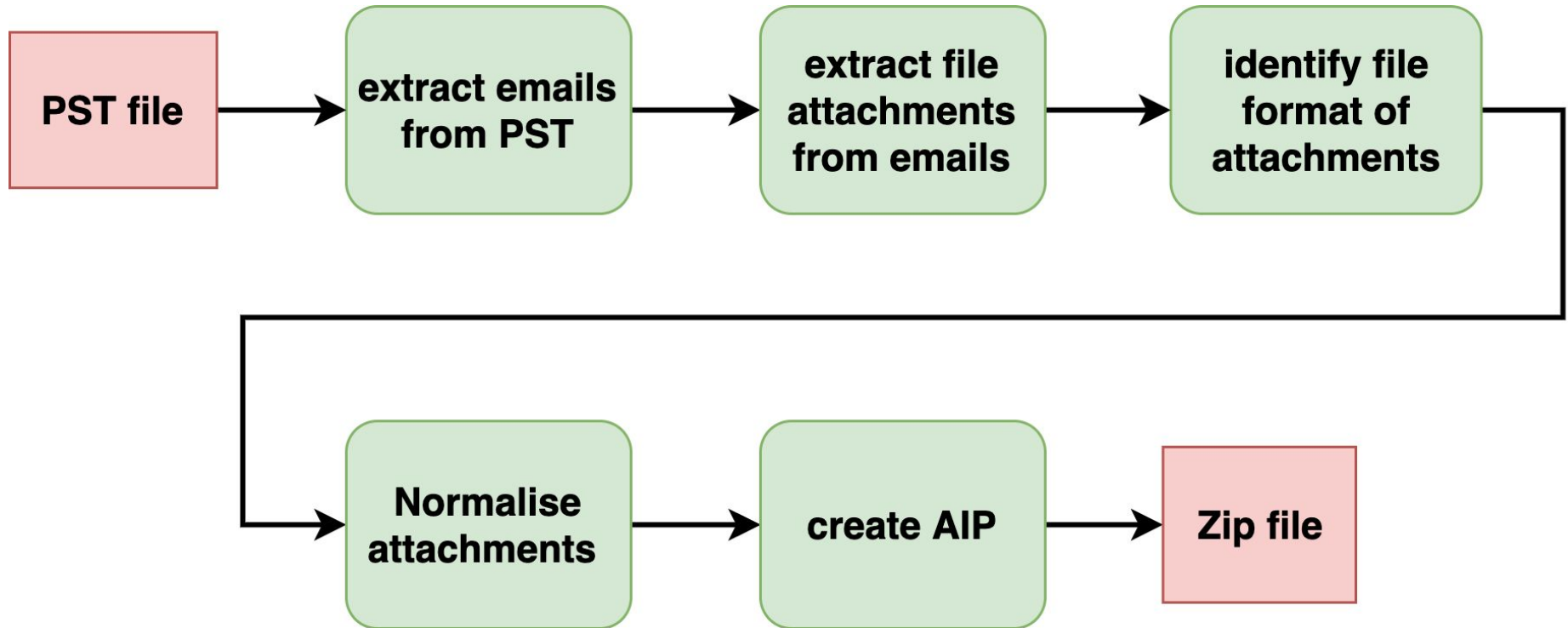
Tool

Property	Value
Name	7Zip
Example	7z x -bd -o"%outputDirectory%" "%inputFile%"
Tool Version	9.20

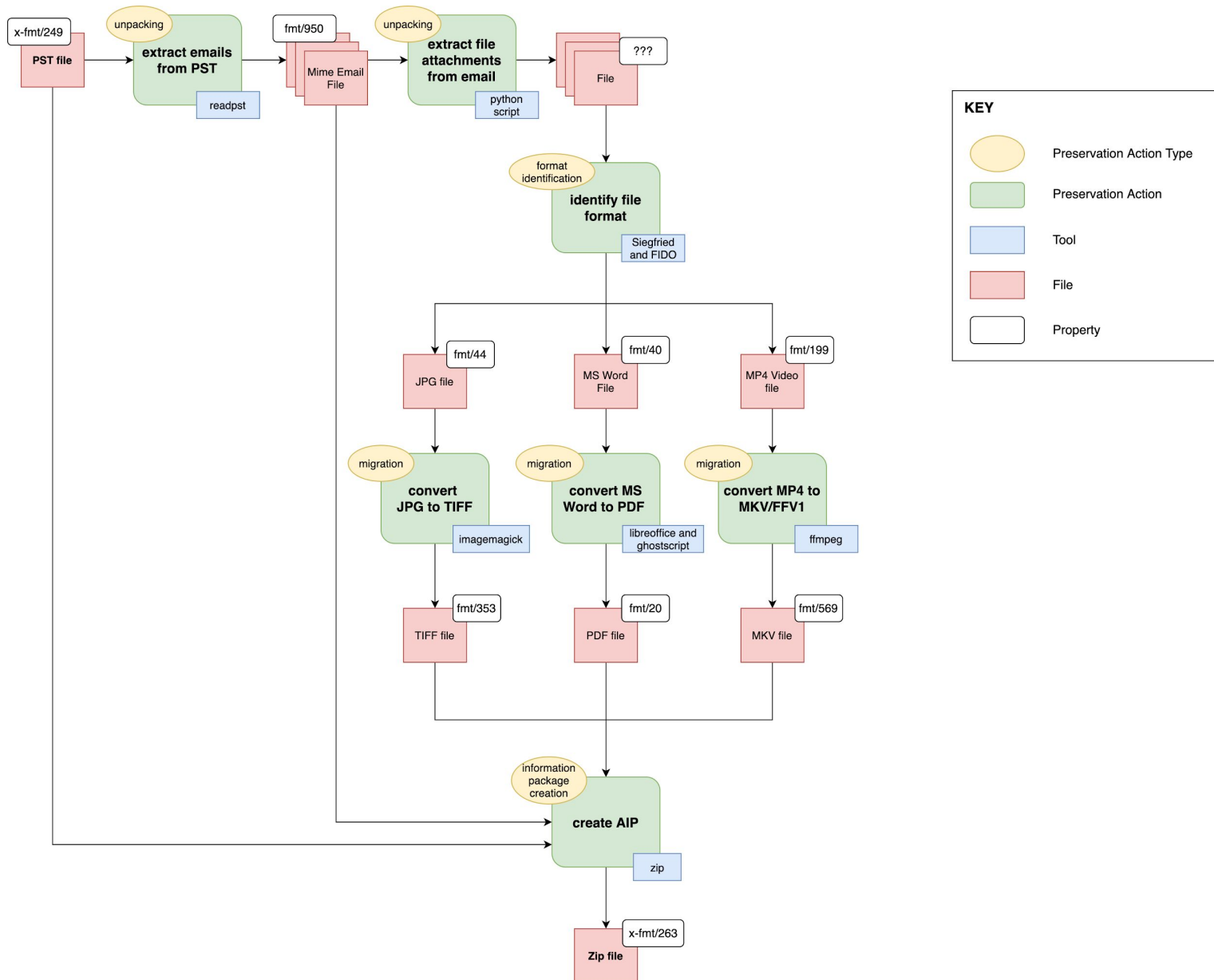
Input

Property	Value
[File input #0] Input File Description (0)	The file that will be acted upon
[File input #0] Input File Name (0)	%inputFile

Email Preservation Workflow using Archivematica



Email workflow expressed as PAR actions



Example PAR Preservation Actions

PST extraction into folders of EML emails

<http://54.77.193.142:5000/api/par/preservation-actions/15adbb46-4e22-4fc5-8376-972d6186000b>

Extraction of file attachments from EML emails

<http://54.77.193.142:5000/api/par/preservation-actions/5b620716-b7d2-439f-b8d8-f6e3c7c5467e>

Normalisation of MS Word attachment to PDF/A

<http://54.77.193.142:5000/api/par/preservation-actions/9ed80b78-b8f1-40a6-8912-cee185c9be6d>



<https://github.com/artefactual-labs/fpr2par>

Preservation Actions in Arkivum Perpetua

[Home](#)

Preservation Action Details

[Setup](#)

Property	Value
Name	Convert MS Word DOC to PDF/A for Preservation
ID	9ed80b78-b8f1-40a6-8912-cee185c9be6d
Namespace	https://archivematica.org
Description	Convert MS Word DOC to PDF/A for Preservation
Last Modified	2020-04-28 07:06:16
Type	normalization

Tool

Property	Value
Name	Arkivum wrapper around LibreOffice and Ghostscript to convert Office formats to PDF/A 2b
Tool Version	Wrapper 0.1.0, LibreOffice 6.0.7.3, ghostscript 9.26

Input

Property	Value
[File input #0] Input File Description (0)	The file that will be acted upon
[File input #0] Input File Name (0)	%inputFile%

Agenda

Item	Presenter
Introduction and Overview	OPF
PAR community stewardship	OPF
PAR use cases	OPF
An overview and demo of the PAR 0.3 release	Jon , Matthew, Jack
Plans for our online workshop	Justin
Q&A	

The PAR workshop

- OPF will host an online interactive workshop on a date to be determined in June 2020
- This will be an opportunity to test out the PAR model and influence its evolution.
- Does not require coding skills or technical work during the workshop
- Workshop Participants will be asked:
 - Can PAR v0.3.0 effectively describe real world preservation scenarios?
 - What changes are needed in the next release of the PAR model to make it more effective, and in what preservation scenarios are they needed?
 - How easy is it to use the PAR model in practice and how could it be simplified or improved?

PAR workshop agenda

PAR Model Overview

- Description of the core entities defined in the PAR Data Model

PAR by Example

- Review of a set of ready-worked examples describing a range of preservation scenarios
- Discussion of changes or enhancements to improve the examples

PAR Roadmap Review

- Review and discussion of changes proposed for PAR v0.4.0

Feedback and next steps

- Discussion and review of next steps

Preparing for the PAR workshop

Example Preservation Action

Description: Extraction of properties for Video files using MediaInfo

Preservation Action Type: metadata extraction

Tool: mediainfo

Tool Version: 18.03

Example: commandline 'mediainfo --Output=EBUCore inputfile

Constraints: allowedFormats (x-fmt/384 QuickTime)

Inputs: inputfile

Outputs: EBUCore XML output from MediaInfo

ID: 1caa0cde-e345-44ac-8d83-51afaa7427b6

Source:

<https://github.com/artefactual-labs/rdss-par/blob/am-characterize-1/examples/preservationAction/mediainfo3.json>

Preparing for the PAR workshop

Example Business Rule

Description: EBU Core XML is the required format for recording technical characteristics of Quicktime Movie Files.

Notes: Modern Institution has decided to adopt EBUCore as the primary metadata schema for describing technical characteristics of Video Files. Only EBUCore metadata is allowed. See https://modern.institute.edu/preservation_policies

Preservation Action Types: metadata extraction

Preservation Actions: MediaInfo3 (1caa0cde-e345-44ac-8d83-51afaa7427b6)

Priority: 1

rawOutputsRetrieved: EBUCore XML output from MediaInfo

Outputs: EBUCore XML output from MediaInfo

ID: 7a71b0d2-d6f3-5d49-a132-9b6776ec6243

Agenda

Item	Presenter
Introduction and Overview	OPF
PAR community stewardship	OPF
PAR use cases	OPF
An overview and demo of the PAR 0.3 release	Jon , Matthew, Jack
Plans for our online workshop	Justin
Q&A	

Q&A

Questions?