



# Preservation Action Registries Workshop

DPC / JISC Workshop 08/07/2019

JISC, Arkivum, Artefactual, Preservica, Open Preservation Foundation

<http://parcore.org/presentations>

V03 140619

# Agenda

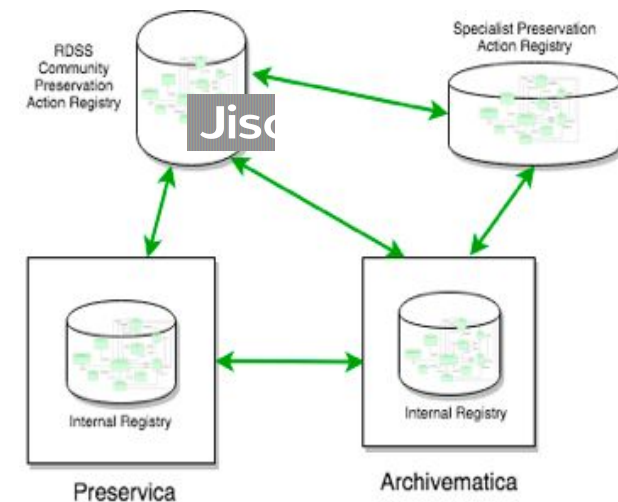
| Time                 | Session                                   | Facilitator/Speaker                                     |
|----------------------|---|---|
| 10:00 - 10:30        | Coffee and registration                   | --  |
| <b>10:30 - 10:40</b> | <b>Introduction / Overview of the day</b> | <b>Martin Wrigley, OPF</b>                              |
| 10:40 - 10:50        | PAR Background and overview               | Matthew Addis, Arkivum                                  |
| 10:50 - 11:10        | PAR Model / Examples                      | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| 11:10 - 11:20        | PAR demo                                  | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| 11:20 - 11:35        | PAR Going Forward / Sustainability        | Carl Wilson / Martin Wrigley, OPF                       |
| 11:35 - 12:00        | Workshop overview                         | Justin Simpson, Artefactual                             |
| 12:00 - 13:00        | Break                                     |   |
| 13:00 - 15:30        | PAR Action Rules Workshop                 | Justin Simpson / All                                    |
| 15:30 - 16:00        | Feedback, Q&A                             | All   |

# Introduction & Overview

- Welcome!
- The purpose of today is to introduce you to the concept of PAR and the work undertaken to date:
  - To get you involved via a collaborative workshop using the PAR model
  - Provide an overview of the vision for PAR and to get your feedback to help refine that vision
  - Introduce the challenges facing PAR and its sustainability
- Quick Fire round table intro:
  - What are your expectations of the day?

# What is PAR about

- PAR is aimed at:
  - a way of describing a digital preservation process that can be shared between preservation systems
  - a way of setting up new systems or processes based on good practice from another, trusted organisation



FPR

# PAR Positioning

- **PAR**
  - Original idea from Jisc Open Research Hub project
    - Multi-vendor shared services platform for Research Data Management
  - Jisc kick-started several systems vendors co-operating to build an interchange of digital preservation policies implemented through specific use of tools and settings
  - Open Research Hub could then build good practice for researchers to use - 'off the shelf' to preserve their data using whatever system their organisation may have.
- What does this translate to in other environments?
- Can we share good practice in this way using PAR?
- Can we make PAR a sustainable success?

# The PAR Team

- Justin Simpson - Managing Director; Artefactual
- Sarah Romkey - Archivemata Program Manager; Artefactual
- Matthew Addis - CTO; Arkivum
- Jack O'Sullivan - Senior Software Engineer; Preservica
- Carl Wilson - Technical Lead; OPF
- Becky McGuinness - Community Manager; OPF
- Martin Speller - Project Manager; OPF
- Martin Wrigley - Executive Director; OPF

# Agenda

| Time                 | Session                            | Facilitator/Speaker                                     |
|----------------------|------------------------------------|---|
| 10:00 - 10:30        | Coffee and registration            | --  |
| 10:30 - 10:40        | Introduction / Overview of the day | Martin Wrigley, OPF                                     |
| <b>10:40 - 10:50</b> | <b>PAR Background and overview</b> | <b>Matthew Addis. Arkivum</b>                           |
| 10:50 - 11:10        | PAR Model / Examples               | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| 11:10 - 11:20        | PAR demo                           | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| 11:20 - 11:35        | PAR Going Forward / Sustainability | Carl Wilson / Martin Wrigley, OPF                       |
| 11:35 - 12:00        | Workshop overview                  | Justin Simpson, Artefactual                             |
| 12:00 - 13:00        | Break                              |   |
| 13:00 - 15:30        | PAR Action Rules Workshop          | Justin Simpson / All                                    |
| 15:30 - 16:00        | Feedback, Q&A                      | All   |

# What is PAR?

- A way to describe **how to do digital preservation**
  - Human and machine readable
  - Enough detail to actually do something in the real world
  - Knowledge sharing through a common language and terminology
- A way to build **registries of good preservation practice**
  - Within an institution, e.g. implementing policies
  - Supporting a community, e.g. research data management
  - Discipline specific, e.g. how to preserve AV content
- A way to achieve **interoperability of DP systems**
  - Import and export from registries and exchange between systems
  - Supports transparency and trust
  - Allows comparison and migration

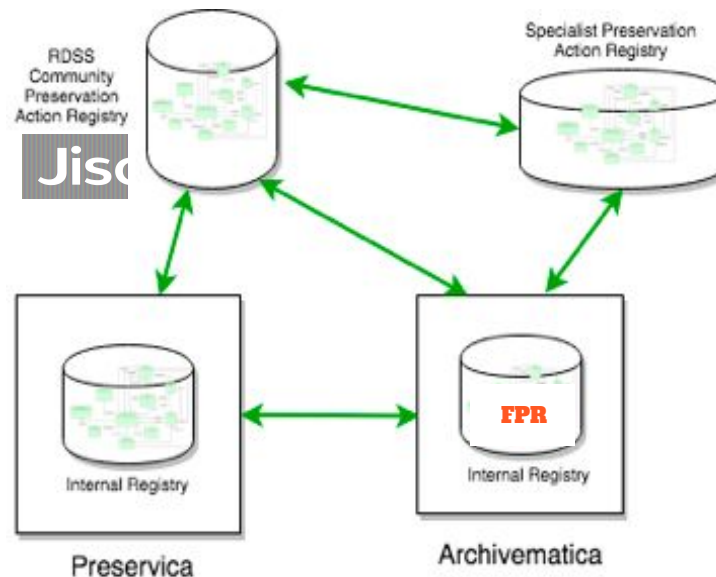


# Why do we need PAR?

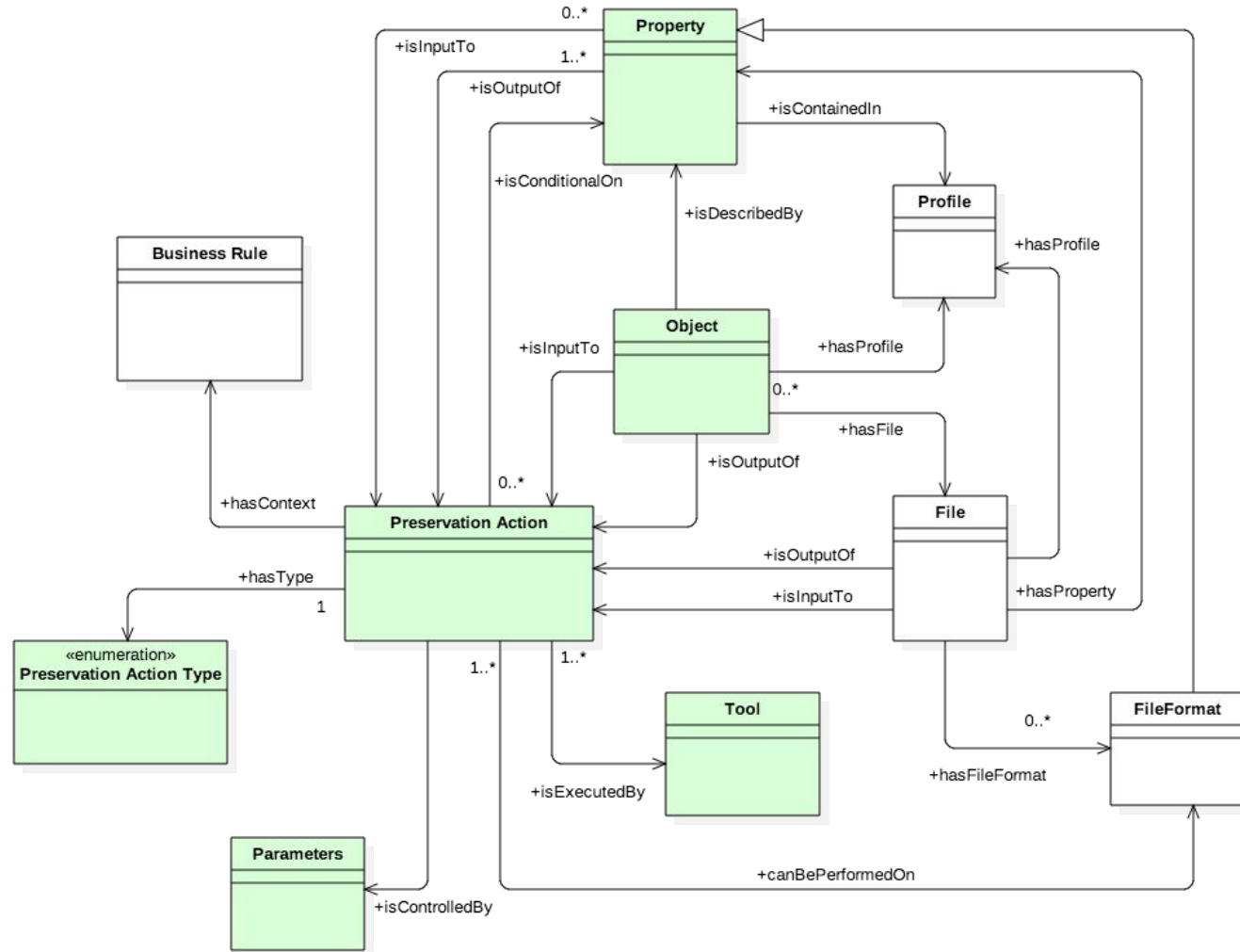
- Users want good advice, wherever it comes from
  - Identification, property extraction, validation, migration, rendering, emulation, packaging, safeguarding ...
  - What works, what doesn't, when to use a particular approach, who's done it before, why they did it ...
- There is plenty of research and advice on good practice
  - Vendors, practitioners, academics, specialists
  - Tools, policies, examples, case studies
- But all this goodness is really hard to find and use
  - Fragmented, hard to find, inconsistent terminology, can't be used directly in DP systems, not always easy to establish trust, lack of context

# Where did PAR come from?

- Jisc Open Research Hub
  - Multi-vendor shared services platform for Research Data Management
- Discussions of interoperability between the DP solutions
  - No common format policies, hard for users to decide which one to use



# What sort of things does PAR describe?



<https://doi.org/10.6084/m9.figshare.6628418>

# Formalising PAR as a set of core concepts

- A Preservation **Action** is something **done** as part of DP
- An **Action** has an **Action Type** defined by PREMIS
- An **Action** acts upon an input **Object** or **File**
- May take **Properties** as inputs
- Executed using one or more **Tools**
- Controlled/configured by a set of **Parameters**
- May create an output **Object** or **File**
- May create/extract **Properties** and provide them as outputs
- Context on when/why/who/how is in **Business Rules**

# Some examples of Preservation Actions

- A **Preservation Actions** is *one or more steps* that are *executed* as part of performing *digital preservation* on *digital content*
- Some simple examples:
  - Create or check checksums
  - Convert a file from format A to format B
  - Extract properties X, Y, Z from a file
  - Validate a file against a format specification
  - Create an Archive Information Package
  - Make replicas and store in different locations

# That doesn't sound too hard...

- Some more complicated examples (beyond files):
  - Extract properties of content in wrappers and containers, e.g. zip, MXF, PST
  - Render compound objects, e.g. digitised books, websites, datasets
  - Migrate structured data, e.g. databases, spreadsheets
  - Render interactive content, e.g. lab notebooks, games, eBooks, installation art
  - Capture shared content and experiences, e.g. multiplayer games, SM
- Preservation Actions on Objects:
  - Often require multiple steps, i.e. they will typically be workflows
  - Need to be done consistently across an intellectual entity as a whole
  - May require preservation of software as well as content

We're starting with the simple stuff and working our way up!

# PAR in the real world (property extraction)

- Property extraction
  - E.g. MediaInfo using EBUCore to AV properties
  - Same Tool, same File, multiple Properties
  - Parameter changes the schema, units, encoding and reported properties!

```
Video
ID : 1
Format : AVC
Format/Info : Advanced Video Codec
Format profile : Main@L3
Format settings : CABAC / 3 Ref Frames
Format settings, CABAC : Yes
Format settings, ReFrames : 3 frames
Codec ID : avc1
Codec ID/Info : Advanced Video Coding
Duration : 13 s 139 ms
Bit rate : 385 kb/s
Width : 854 pixels
Height : 480 pixels
Display aspect ratio : 16:9
Frame rate mode : Variable
Frame rate : 23.976 FPS
Minimum frame rate : 23.974 FPS
Maximum frame rate : 23.981 FPS
Color space : YUV
Chroma subsampling : 4:2:0
Bit depth : 8 bits
```

```
<ebucore:videoFormat videoFormatName="AVC">
  <ebucore:width unit="pixel">854</ebucore:width>
  <ebucore:height unit="pixel">480</ebucore:height>
  <ebucore:frameRate factorNumerator="999" factorDenominator="1000">24</ebucore:frameRate>
  <ebucore:aspectRatio typeLabel="display">
    <ebucore:factorNumerator>16</ebucore:factorNumerator>
    <ebucore:factorDenominator>9</ebucore:factorDenominator>
  </ebucore:aspectRatio>
  <ebucore:videoEncoding typeLabel="Main@L3"/>
  <ebucore:codec>
    <ebucore:codecIdentifier>
      <dc:identifier>avc1</dc:identifier>
    </ebucore:codecIdentifier>
  </ebucore:codec>
  <ebucore:bitRate>385204</ebucore:bitRate>
```

mediainfo video.mp4

mediainfo --Output=EBUCore video.mp4

# PAR in the real world (fixity)

- Checksum generation and validation
  - md5sum can generate an MD5 checksum for a file
  - md5sum can check an MD5 to get OK/FAIL for a file
  - Same Tool, same File, same Property
  - A Parameter changes the Action and Action Type (fixity check, digest calc)
  - A Fixity check is Property extraction followed by Property comparison

What is the fixity of parcore.ppt?

md5sum parcore.ppt: 92bc215089ccf35a8384fc25f9be1bd3 parcore.ppt

md5sum -c manifest.md5: parcore.ppt: OK



# PAR in the real world (file format identification)

- File format identification
  - Same Action, multiple Tools: Droid, Siegfried, Fido, FITS, File, Tika
  - Can be applied to wrappers, containers, bitstreams
  - Different schemas: Mimetype, PRONOM ID
  - Business Rules needed on which Tools work best and when
  - Business Rules needed on priority/ranking/consensus/conflict resolution

What is the file format of parcore.ppt?

|                   |  |
|-------------------|--|
| File:             | Composite Document File V2 Document          |
| Fido:             | fmt/111, “OLE2 Compound Document Format”     |
| Siegfried, DROID: | fmt/126, “Microsoft Powerpoint Presentation” |
| Tika:             | Content-Type: application/vnd.ms-powerpoint  |

# How much of PAR currently exists?

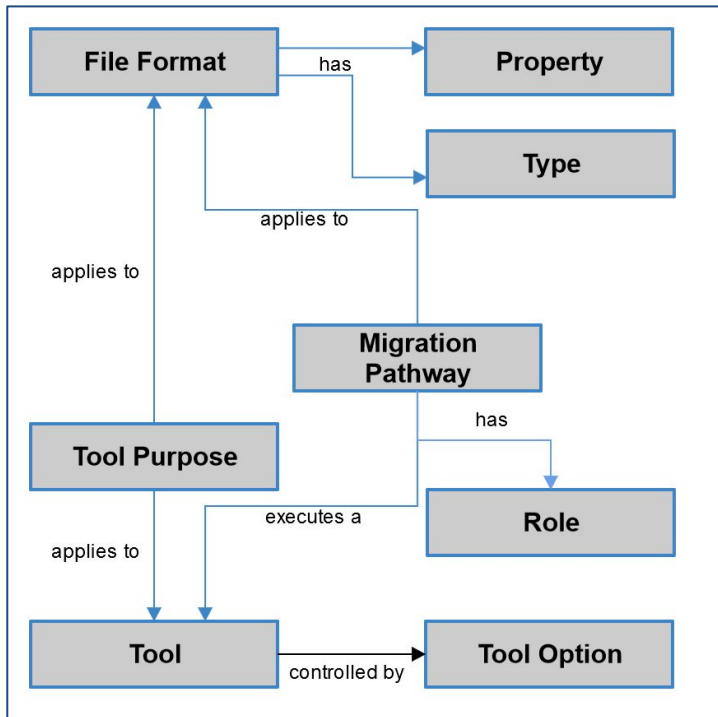
|                       |  |
|-----------------------|--|
| Glossary              | <ul style="list-style-type: none"><li>● Definition of the PAR core concepts</li></ul>  |
| Conceptual Model      | <ul style="list-style-type: none"><li>● Common framework for everyone to work to</li><li>● Something to argue about and agree on!</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>                       |

# Agenda

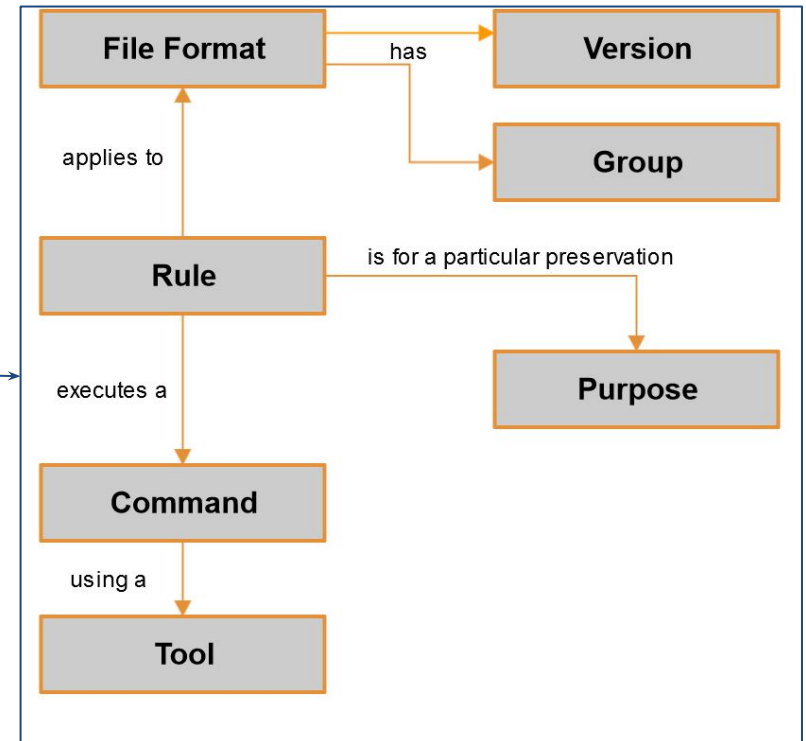
| Time                 | Session                            | Facilitator/Speaker  |
|----------------------|------------------------------------|--|
| 10:00 - 10:30        | Coffee and registration            | --   |
| 10:30 - 10:40        | Introduction / Overview of the day | Martin Wrigley, OPF  |
| 10:40 - 10:50        | PAR Background and overview        | Matthew Addis, Arkivum   |
| <b>10:50 - 11:10</b> | <b>PAR Model / Examples</b>        | <b>Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual</b> |
| 11:10 - 11:20        | PAR demo                           | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual        |
| 11:20 - 11:35        | PAR Going Forward / Sustainability | Carl Wilson / Martin Wrigley, OPF                              |
| 11:35 - 12:00        | Workshop overview                  | Justin Simpson, Artefactual                                    |
| 12:00 - 13:00        | Break                              |  |
| 13:00 - 15:30        | PAR Action Rules Workshop          | Justin Simpson / All   |
| 15:30 - 16:00        | Feedback, Q&A                      | All  |

# Registry (In)compatibility

## Preservica Registry

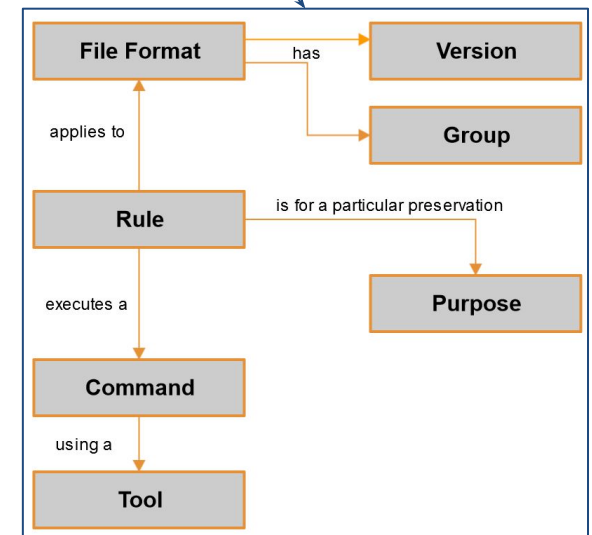
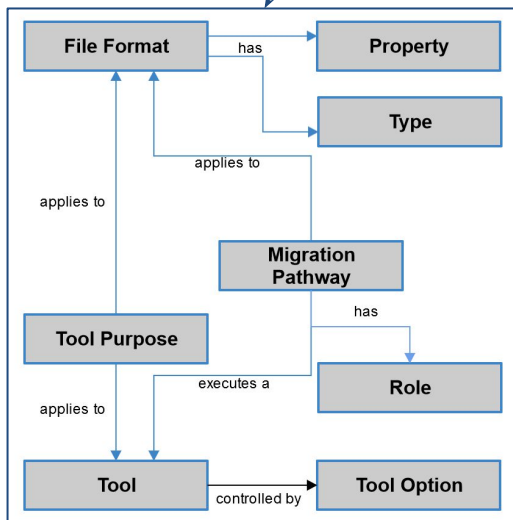
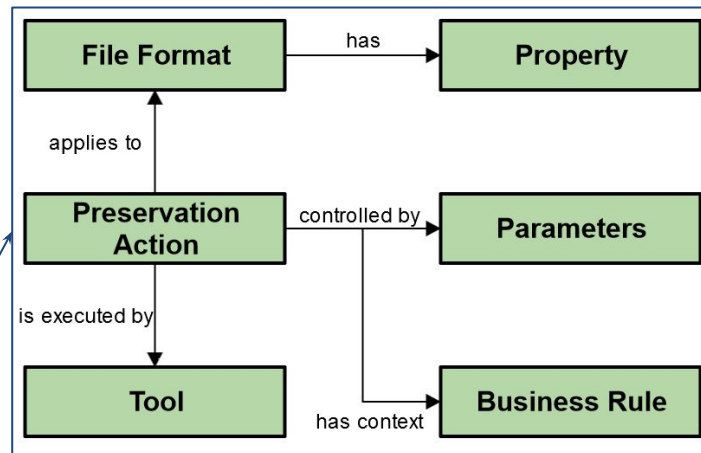


## Archivematica FPR



?

# Common Language



# Extract Metadata Use Case

- User Bob at Modern Institution is responsible for preserving AV content.
- Modern Institution has decided to adopt EBUcore as the metadata standard for its AV holdings.
- Bob discovers that a recent update to the MediaInfo tool allows him to extract metadata about AV files in EBUCore XML format.
- He introduces this change into his Preservica Preservation System and exports this information as a Preservation Action expressed in PAR Schema.

# Extract Metadata Use Case

- User Alice at Post-Modern Institution imports this newly published information into her Archivemata preservation system using its PAR API.
- Each PAR entry is stored in her installation as JSON files.
- She does not have to read it in JSON or even PAR schema format or have any knowledge about how Bob's system works.
- She instead reviews it in a PAR Form that appears under Archivemata's Preservation Planning tab.

# Extract Metadata Use Case

- Alice decides that this new capability is something she wants to implement.
- She selects the 'Convert to FPR' option to enter it as an active 'Metadata Extraction' Rule and Command in her Archivematica instance.
- At this point the PAR information is entered into the Archivematica database without affecting any existing Commands or Rules.
- The new Rules and Command are linked to the PAR Preservation Action.



# Agenda

| Time                 | Session                            | Facilitator/Speaker  |
|----------------------|------------------------------------|--|
| 10:00 - 10:30        | Coffee and registration            | --   |
| 10:30 - 10:40        | Introduction / Overview of the day | Martin Wrigley, OPF  |
| 10:40 - 10:50        | PAR Background and overview        | Matthew Addis, Arkivum   |
| 10:50 - 11:10        | PAR Model / Examples               | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual        |
| <b>11:10 - 11:20</b> | <b>PAR demo</b>                    | <b>Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual</b> |
| 11:20 - 11:35        | PAR Going Forward / Sustainability | Carl Wilson / Martin Wrigley, OPF                              |
| 11:35 - 12:00        | Workshop overview                  | Justin Simpson, Artefactual                                    |
| 12:00 - 13:00        | Break                              |  |
| 13:00 - 15:30        | PAR Action Rules Workshop          | Justin Simpson / All   |
| 15:30 - 16:00        | Feedback, Q&A                      | All  |

# PAR Demo

- Demo Export and Import of Preservation Action

<https://youtu.be/ELjbgksjBmU>

# Agenda

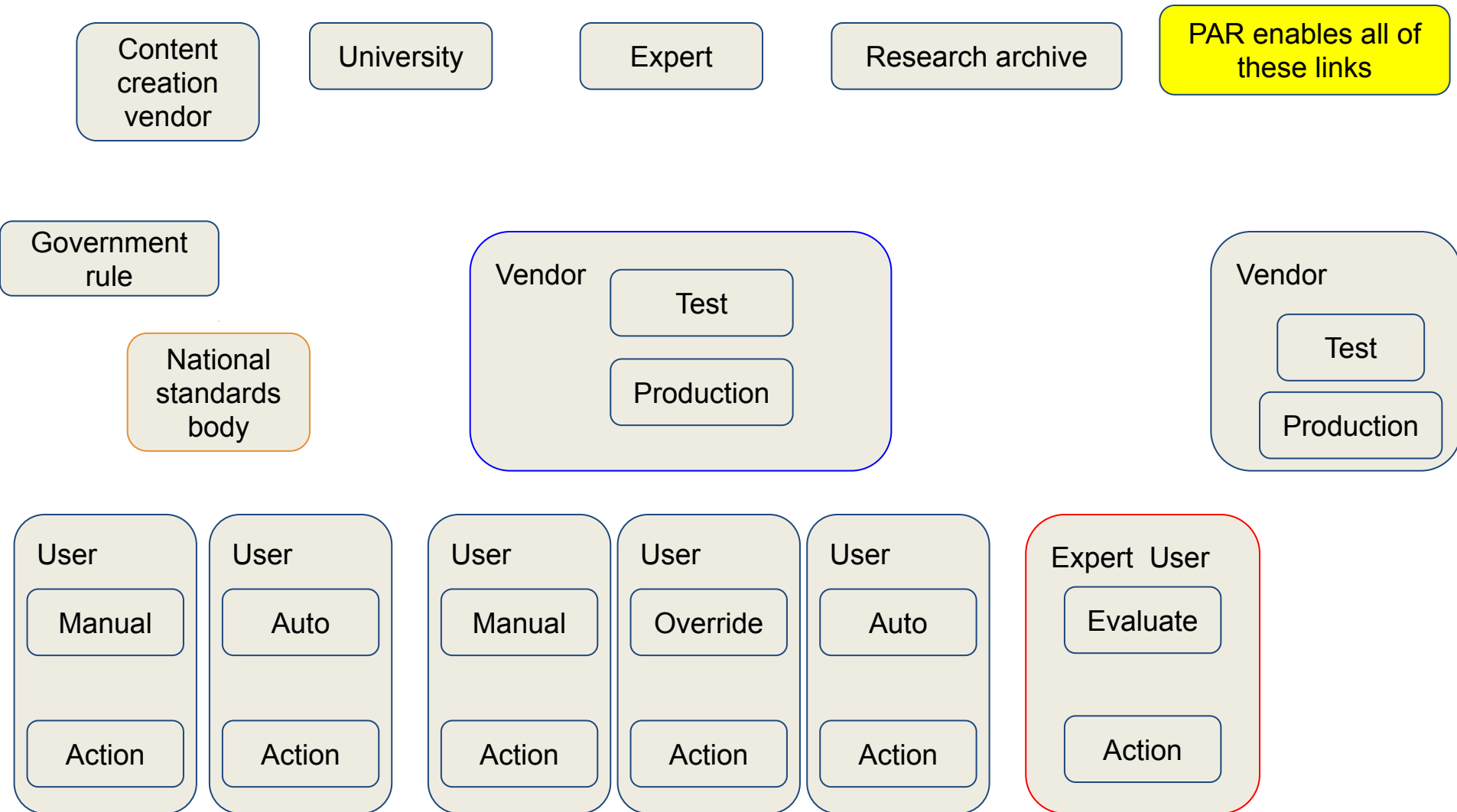
| Time                 | Session                                   | Facilitator/Speaker                                     |
|----------------------|---|---|
| 10:00 - 10:30        | Coffee and registration                   | --  |
| 10:30 - 10:40        | Introduction / Overview of the day        | Martin Wrigley, OPF                                     |
| 10:40 - 10:50        | PAR Background and overview               | Matthew Addis, Arkivum                                  |
| 10:50 - 11:10        | PAR Model / Examples                      | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| 11:10 - 11:20        | PAR demo                                  | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| <b>11:20 - 11:35</b> | <b>PAR Going Forward / Sustainability</b> | <b>Carl Wilson / Martin Wrigley, OPF</b>                |
| 11:35 - 12:00        | Workshop overview                         | Justin Simpson, Artefactual                             |
| 12:00 - 13:00        | Break                                     |   |
| 13:00 - 15:30        | PAR Action Rules Workshop                 | Justin Simpson / All                                    |
| 15:30 - 16:00        | Feedback, Q&A                             | All   |

# Long Term Vision

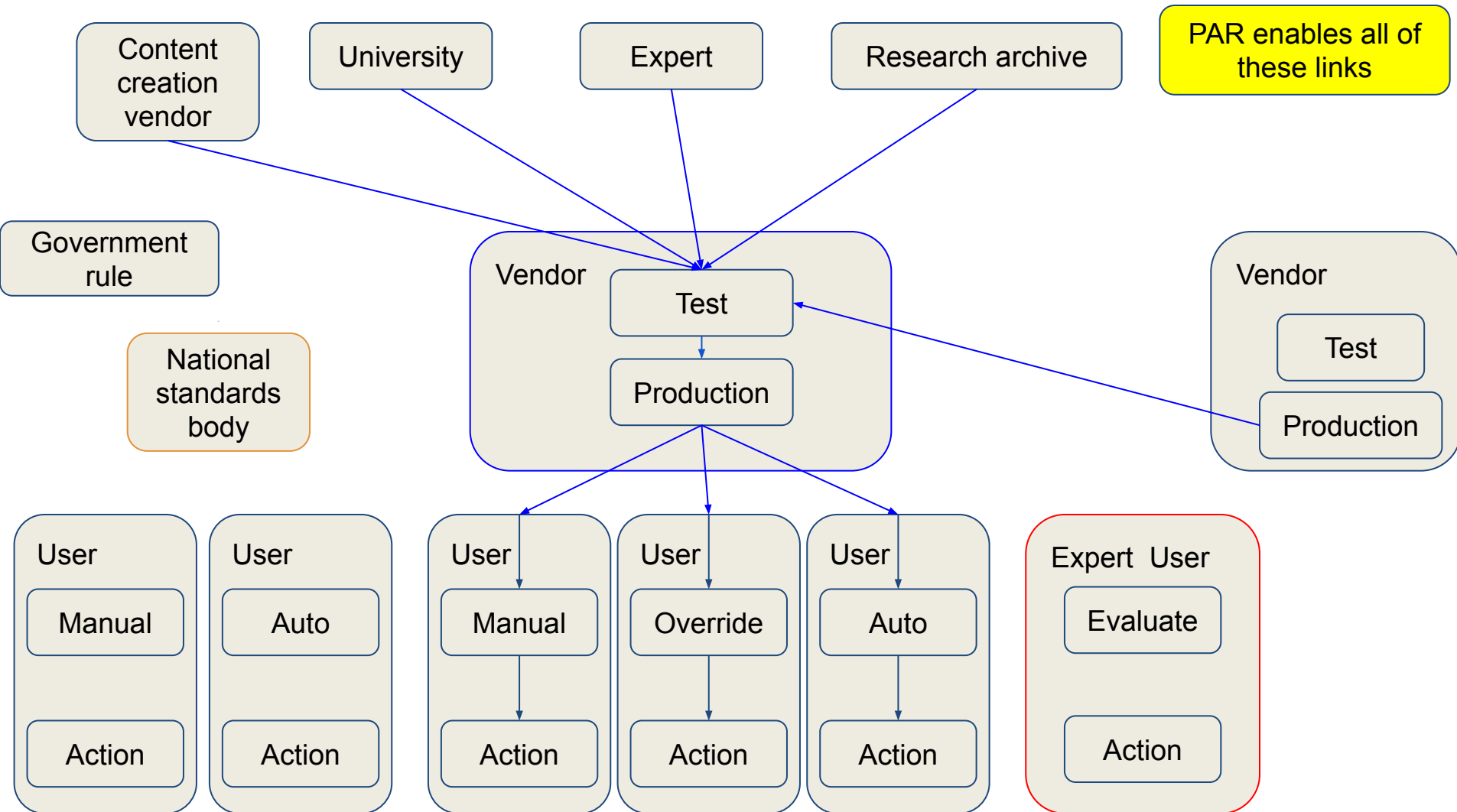
- All users want trusted information
- Some users want to make their own choices
- Some countries or communities have local rules for preservation formats
- Recommended action depends on user context e.g. public access vs minimum storage cost
- Many users want it to just work for their product in their context

*PAR must be able to help in all of these situations*

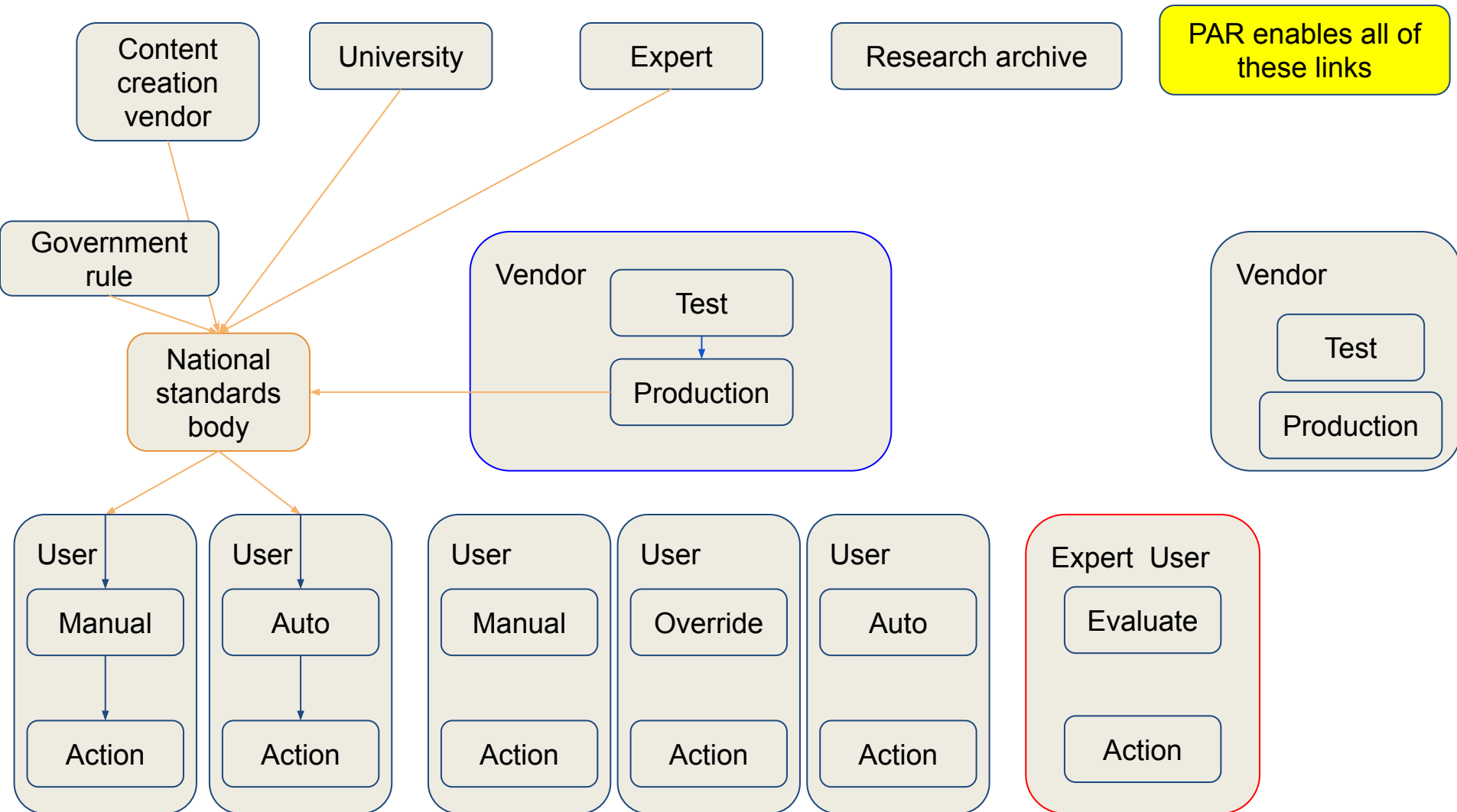
# Long Term Vision



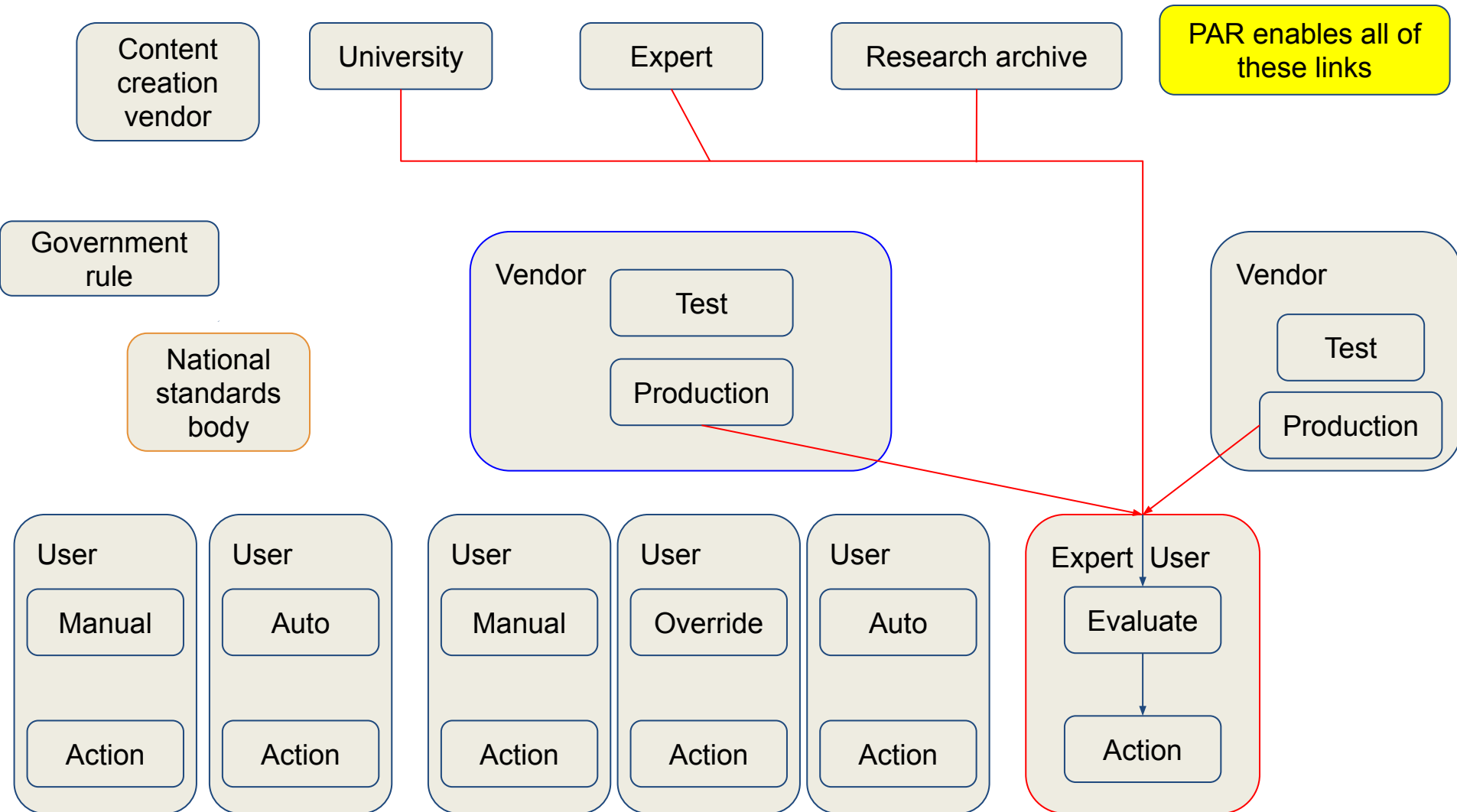
# Long Term Vision



# Long Term Vision



# Long Term Vision





# Proposed PAR Development Plan

- To-date PAR has delivered a Proof of Concept largely funded by JISC as part of the RDSS project
- Working with the DP Community to gather user examples and adding local registry endpoints
- A plan has been created to drive PAR forward taking it to a beta production state
  - the required effort is estimated at 1 person-year

# Proposed PAR Development Plan

- The plan includes a set of development sprints aimed at:
  - Creating a more robust data model, business rule set and supporting specifications
  - Building out the detailed PAR model
  - Enhancing the API to handle complex payloads
  - Development of a Sync tool to align / compare registries
  - Updates to proprietary systems

# PAR Sustainability

- The initial round of PAR funding is coming to an end
  - (Many thanks to Jisc for starting this all off!)
- The PAR core team continues to collaborate on a regular basis defining the detailed work scope associated with 'productising' PAR
- The next step is to work closely with the DP Community to identify a viable future for PAR - based on a clear commercial need - and then to explore how we might secure any further funding that might be required

# Agenda

| Time                 | Session                            | Facilitator/Speaker                                     |
|----------------------|------------------------------------|---|
| 10:00 - 10:30        | Coffee and registration            | --  |
| 10:30 - 10:40        | Introduction / Overview of the day | Martin Wrigley, OPF                                     |
| 10:40 - 10:50        | PAR Background and overview        | Matthew Addis, Arkivum                                  |
| 10:50 - 11:10        | PAR Model / Examples               | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| 11:10 - 11:20        | PAR demo                           | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| 11:20 - 11:35        | PAR Going Forward / Sustainability | Carl Wilson / Martin Wrigley, OPF                       |
| <b>11:35 - 12:00</b> | <b>Workshop overview</b>           | <b>Justin Simpson, Artefactual</b>                      |
| 12:00 - 13:00        | Break                              |   |
| 13:00 - 15:30        | PAR Action Rules Workshop          | Justin Simpson / All                                    |
| 15:30 - 16:00        | Feedback, Q&A                      | All   |

# PAR Workshop Overview

- Break into groups
- See one
  - review examples
- Do one
  - Write down new examples as a group
    - Business Rule
    - Preservation Action
- Teach one
  - Pair up with someone from another group
    - Explain your groups work
    - Listen to their explanation of their groups work

# PAR Workshop Overview

## Business Rules

- Description
- Notes
- Preservation Action Types
- Formats
- Format Families
- **ID**
- Preservation Actions
  - Priority
  - Inputs
  - Outputs

## Preservation Actions

- **Description**
- **Preservation Action Type**
- **Tool**
- Tool Version
- Example
- Constraints
- Inputs
- Outputs
- **ID**

# PAR Workshop Overview

## Business Rules

- Description
- Notes
- Preservation Action Types
- Formats
- Format Families
- **ID**
- Preservation Actions
  - Priority
  - Inputs
  - Outputs

Business Rules explain why particular Preservation Actions are performed in particular contexts

Mostly narrative, but they provide a way to

- a) explicitly link Preservation Actions to policies
- b) Prioritize Preservation Actions
- c) Specify required inputs
- d) Specify expected outputs

# PAR Workshop Overview

## Business Rules Details

- Description
  - *A short human readable explanation/display name for the Business Rule*
- Notes
  - *A free text field for providing additional context. This may be used to record the decision making process that led to the formulation of this rule, details of real-world experience in applying the rule, or any other text.*
- Preservation Action Types
  - *One or more LoC Preservation Event Types*
  - *A list of Preservation Action Types that this Business Rule should be applied to. This might be a subset of those that the Preservation Actions themselves apply to.*
- Format Families
  - *A list of format families that this Business Rule should be applied to*
- Formats
  - *A list of file formats that this Business Rule should be applied to*



# PAR Workshop Overview

## Business Rules Details

- Preservation Actions
  - *A list of specific actions. These are defined internally to reference the Core Preservation Action, the priority order in which it should be performed, and any inputs and outputs that should be used.*

### Actions

- Preservation action
  - Optional input properties
  - Output files retrieved
  - Output properties retrieved
  - Raw outputs retrieved
- Priority

# PAR Workshop Overview

## Preservation Actions

- **Description**
- **Preservation Action Type**
- **Tool**
- Tool Version
- Example
- Constraints
- Inputs
- Outputs
- **ID**

Preservation Actions are processes that are run as part of performing digital preservation.

- Classified by Preservation Action Type
- Executed by one or more Tools
- Constraints define when to use this action
- Inputs define files or properties required to execute the action
- Outputs define objects or properties created by the action

# PAR Workshop Overview

## Preservation Action Details

- Description
  - *A short human readable explanation/display name*
- Type
  - *One LoC Preservation Event Type*
  - *A Preservation Action Type that this Preservation Action performs.*
- Tool
  - *A PAR Tool*
- Tool Version
  - *The specific version of the tool this Action requires*
- Example
  - *A human readable explanation of how to execute the Preservation Action.*

# PAR Workshop Overview

## Preservation Action Details

- Constraints
  - *Defines limitations of when to perform this Action*
    - Limit to files of specific formats, or
    - Limit to objects with specific properties
  - AllowedFormats, AllowedPropertiesAnyOf, AllowedPropertiesAllOf
- Inputs
  - *A list of expected inputs (files or properties)*
  - File, Property
- Outputs
  - *A list of outputs created by this Action*
  - File, Property, Raw

# Agenda

| Time                 | Session                            | Facilitator/Speaker                                     |
|----------------------|------------------------------------|---|
| 10:00 - 10:30        | Coffee and registration            | --  |
| 10:30 - 10:40        | Introduction / Overview of the day | Martin Wrigley, OPF                                     |
| 10:40 - 10:50        | PAR Background and overview        | Matthew Addis, Arkivum                                  |
| 10:50 - 11:10        | PAR Model / Examples               | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| 11:10 - 11:20        | PAR demo                           | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| 11:20 - 11:35        | PAR Going Forward / Sustainability | Carl Wilson / Martin Wrigley, OPF                       |
| 11:35 - 12:00        | Workshop overview                  | Justin Simpson, Artefactual                             |
| <b>12:00 - 13:00</b> | <b>Break</b>                       |   |
| 13:00 - 15:30        | PAR Action Rules Workshop          | Justin Simpson / All                                    |
| 15:30 - 16:00        | Feedback, Q&A                      | All   |

# Agenda

| Time                 | Session                            | Facilitator/Speaker                                     |
|----------------------|------------------------------------|---|
| 10:00 - 10:30        | Coffee and registration            | --  |
| 10:30 - 10:40        | Introduction / Overview of the day | Martin Wrigley, OPF                                     |
| 10:40 - 10:50        | PAR Background and overview        | Matthew Addis, Arkivum                                  |
| 10:50 - 11:10        | PAR Model / Examples               | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| 11:10 - 11:20        | PAR demo                           | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| 11:20 - 11:35        | PAR Going Forward / Sustainability | Carl Wilson / Martin Wrigley, OPF                       |
| 11:35 - 12:00        | Workshop overview                  | Justin Simpson, Artefactual                             |
| 12:00 - 13:00        | Break                              |   |
| <b>13:00 - 15:30</b> | <b>PAR Action Rules Workshop</b>   | <b>Justin Simpson / All</b>                             |
| 15:30 - 16:00        | Feedback, Q&A                      | All   |

# PAR Workshop

## Business Rules Details

- Description
  - *A short human readable explanation/display name for the Business Rule*
- Notes
  - *A free text field for providing additional context. This may be used to record the decision making process that led to the formulation of this rule, details of real-world experience in applying the rule, or any other text.*
- Preservation Action Types
  - *One or more LoC Preservation Event Types*
  - *A list of Preservation Action Types that this Business Rule should be applied to. This might be a subset of those that the Preservation Actions themselves apply to.*
- Format Families
  - *A list of format families that this Business Rule should be applied to*
- Formats
  - *A list of file formats that this Business Rule should be applied to*

# PAR Workshop

## Business Rules Details

- Description
  - *A short human readable explanation/display name for the Business Rule*
- Notes
  - *A free text field for providing additional context. This may be used to record the decision making process that led to the formulation of this rule, details of real-world experience in applying the rule, or any other text.*
- Preservation Action Types
  - *One or more LoC Preservation Event Types*
  - *A list of Preservation Action Types that this Business Rule should be applied to. This might be a subset of those that the Preservation Actions themselves apply to.*
- Format Families
  - *A list of format families that this Business Rule should be applied to*
- Formats
  - *A list of file formats that this Business Rule should be applied to*



# PAR Workshop

## Business Rules Details

- Preservation Actions
  - *A list of specific actions. These are defined internally to reference the Core Preservation Action, the priority order in which it should be performed, and any inputs and outputs that should be used.*

### Actions

- Preservation action
  - Optional input properties
  - Output files retrieved
  - Output properties retrieved
  - Raw outputs retrieved
- Priority

# PAR Workshop Overview

**archivematica.** Transfer Backlog Appraisal Ingest Archival storage **Preservation planning** Access Administration test ▾

PAR Preservation Actions / List

**FPR** **PAR**

Business Rules | Preservation Actions

---

---

---

**description** EBU Core XML is the required format for recording technical characteristics of Quicktime Movie files.

**id** e42aa5b2-5652-4a01-a037-7d0ae817a73d

**notes** Modern Institution has decided to adopt EBUCore as the primary metadata schema for describing technical characteristics of Quicktime Video Files. Only EBUCore metadata is allowed. See [http://modern-institution.edu/preservation\\_policies](http://modern-institution.edu/preservation_policies)

**preservationActionTypes** metadata extraction (<http://id.loc.gov/vocabulary/preservation/eventType/mee>)

**preservationActions** MedialInfo3 (1caa0cde-e345-44ac-8d83-51afaa7427b6)

**priority** 1

**rawOutputsRetrieved** EBUCore XML output from MedialInfo

Convert to FPR

---

---

---

# PAR Workshop

## Business Rule example one - Choice of file format ID tools

- **Description:** For File Format Identification of any type of file, use DROID (PAR-tool/1) as the first preference, and FIDO (PAR-tool/2) as the second
- **Notes:** My order of preference/priority when using multiple tools are available for file format identification is DROID then FIDO
- **Preservation Action Types:** format identification
- **Formats:**
- **Format Families:** ALL
- **ID:** a593036d-7427-54aa-b8c0-50a5fb7bd50b
- **Preservation Actions**
  - action/1
    - Priority: 1
    - Inputs
    - Outputs
  - action/2
    - Priority: 2
    - Inputs
    - Outputs

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

# PAR Workshop

## Business Rule example two - Migrating AVI files

- **Description:** FFMPEG is preferable to HandBrake for migrating AVI files to WebM
- **Notes:** We have found that the HandBrake process occasionally hangs when performing this type of migration. FFMPEG displays no similar behaviour and therefore should be used in preference
- **Preservation Action Types:** migration; normalization
- **Formats:** fmt/5
- **Format Families:** AVI
- **ID:** bc6c7498-2b8d-5314-bb0d-c773e72e9ff2
- **Preservation Actions**
  - action/123
    - Priority: 1
    - Inputs
    - Outputs
  - action/456
    - Priority: 2
    - Inputs
    - Outputs

<https://github.com/artefactual-labs/rdss-par/blob/am-characterize-1/examples/br-2.json>

# PAR Workshop

## Business Rule example three - Resizing for access

- **Description:** Use resizing parameters on ImageMagick to perform generation of consistent size access copy JPEGs from TIFFs
- **Notes:** We want all JPEGs for presentation to fit into a 250x250px box so that they display properly in our access system, where the template for displaying images has a 300px div for the image itself, however, we only want to resize where the image itself is larger than that box.
- **Preservation Action Types:** migration
- **Formats:** fmt/353; fmt/155; fmt/154 [etc]
- **Format Families:** TIFF
- **ID:** 7a71b0d2-d6f3-5d49-a132-9b6776ec6243
- **Preservation Actions**
  - action/25
    - Priority: 1
    - Inputs: Resize to fit in 250px square box [etc]
    - Outputs:

<https://github.com/artefactual-labs/rdss-par/blob/am-characterize-1/examples/br-3.json>

# PAR Workshop

## Preservation Action Example 1: Characterize with MediaInfo

- **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`
- **Inputs** `inputfile`
- **Outputs** `EBUCore XML output from MediaInfo`
- **ID** `1caa0cde-e345-44ac-8d83-51afaa7427b6`

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

# PAR Workshop

## Preservation Action Example 2: Validation of checksum with md5sum

- **Description** Validation of an MD5 checksum on a File using md5sum
- **Preservation Action Type** fixity check
- **Tool** md5sum
- **Tool Version**
- **Example** commandline 'md5sum -c manifest.md5', where manifest.md5 contains a MD5 checksum for a file called inputfile
- **Constraints**
- **Inputs** Manifest file containing the MD5 and name of the file to be checked
- **Outputs** Fixity PASS or FAIL
- **ID** f2e953e4-425e-5ed1-a65e-efd0b2e061be

<https://github.com/artefactual-labs/rdss-par/blob/am-characterize-1/examples/md5check1.json>

# Agenda

| Time                 | Session                            | Facilitator/Speaker                                     |
|----------------------|------------------------------------|---|
| 10:00 - 10:30        | Coffee and registration            | --  |
| 10:30 - 10:40        | Introduction / Overview of the day | Martin Wrigley, OPF                                     |
| 10:40 - 10:50        | PAR Background and overview        | Matthew Addis, Arkivum                                  |
| 10:50 - 11:10        | PAR Model / Examples               | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| 11:10 - 11:20        | PAR demo                           | Jack O'Sullivan, Preservica / Sarah Romkey, Artefactual |
| 11:20 - 11:35        | PAR Going Forward / Sustainability | Carl Wilson / Martin Wrigley, OPF                       |
| 11:35 - 12:00        | Workshop overview                  | Justin Simpson, Artefactual                             |
| 12:00 - 13:00        | Break                              |   |
| 13:00 - 15:30        | PAR Action Rules Workshop          | Justin Simpson / All                                    |
| <b>15:30 - 16:00</b> | <b>Feedback, Q&amp;A</b>           | <b>All</b>  |



# 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/>
- Webinar
  - <http://openpreservation.org/event/introducing-preservation-action-registries/> (OPF login required)