Preservation Action Registries

Carl Wilson (OPF), Jon Tilbury (Preservica), Sarah Romkey (Artefactual)

JISC
Arkivum, Artefactual, Preservica, Open Preservation Foundation

http://parcore.org/presentations/

Agenda

- Part 1: Background & Motivation
- Part 2: Project outcomes
- Part 3: Proof of concept implementation
- Part 4: Long Term Vision
- Part 5: How can you get involved

Agenda

- Part 1: Background & Motivation
- Part 2: Project outcomes
- Part 3: Proof of concept implementation
- Part 4: Long Term Vision
- Part 5: How can you get involved

Background: The problem

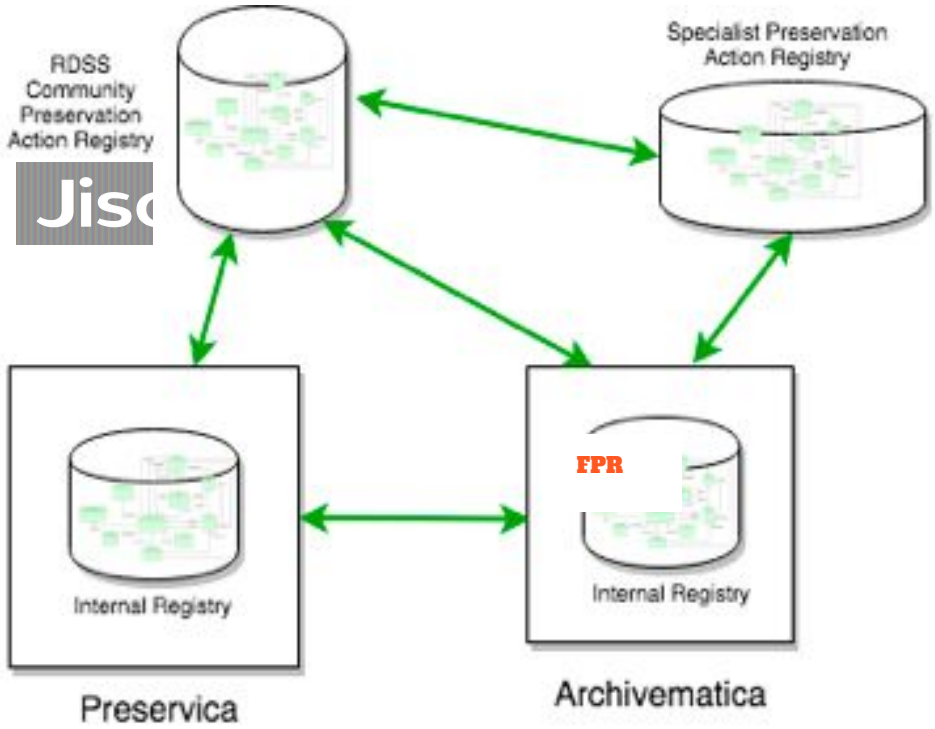
- Users want the best advice, wherever it comes from
  - Identification, property extraction, validation, migration, rendering, tools
- Multiple parallel initiatives research and advise on best practice
  - Products such as Preservica & Archivematica
  - Practitioners
  - Academics
  - Specialists
- 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

Agenda

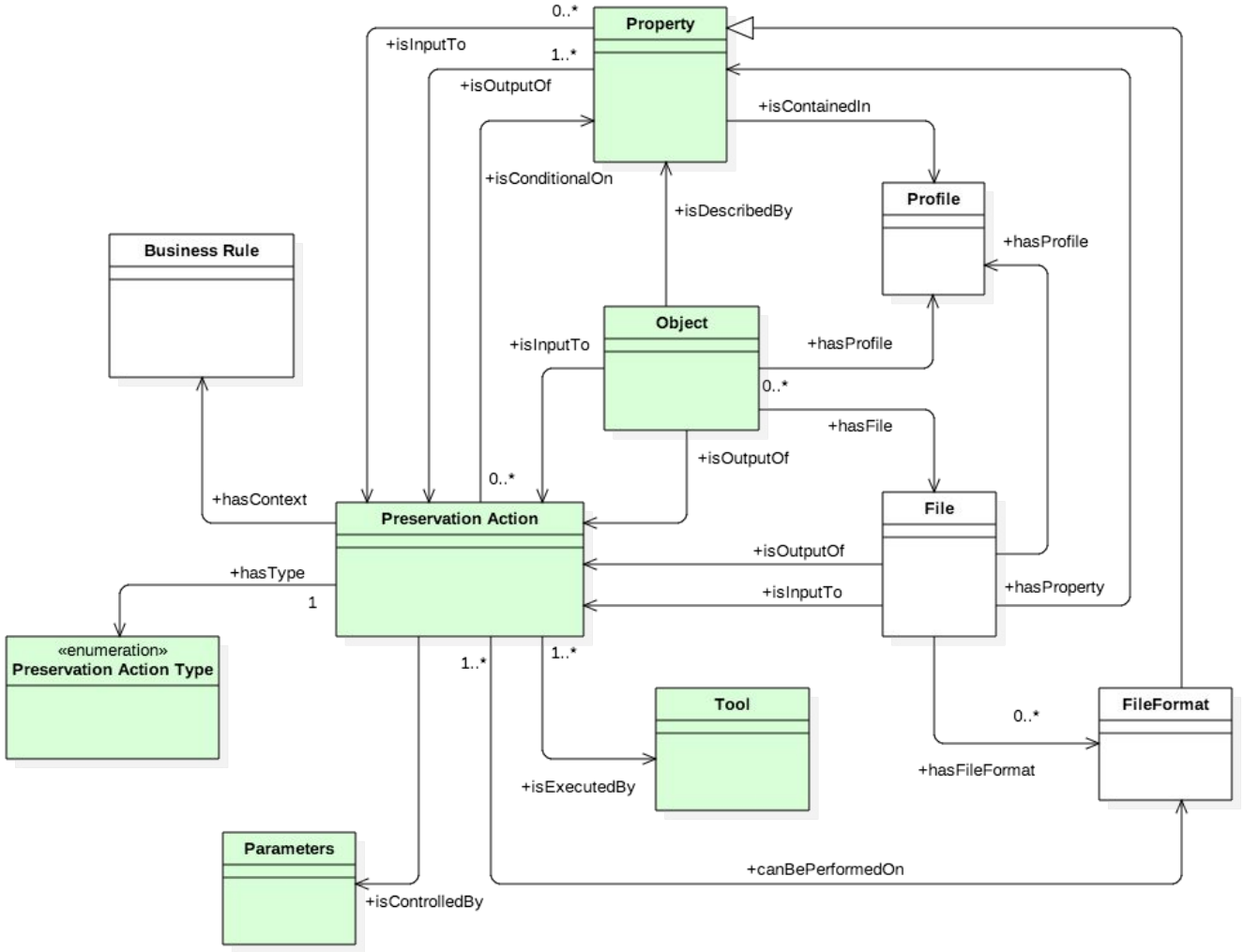
- Part 1: Background & Motivation
- Part 2: Project outcomes
- Part 3: Proof of concept implementation
- Part 4: Long Term Vision
- Part 5: How can you get involved

What have we produced and why?

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

Open Preservation Foundation arkivum Bringing archived data to life Preservica Active digital preservation artefactual Jisc

PAR Conceptual Model



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

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

Open Preservation Foundation

arkivum Bringing archived data to life

Preservica Active digital preservation

artefactual

Jisc

Core concepts: summary

- 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 as outputs
- Context on when/why/who/how is in **Business Rules**

Concrete Examples (why it can get a bit tricky!)

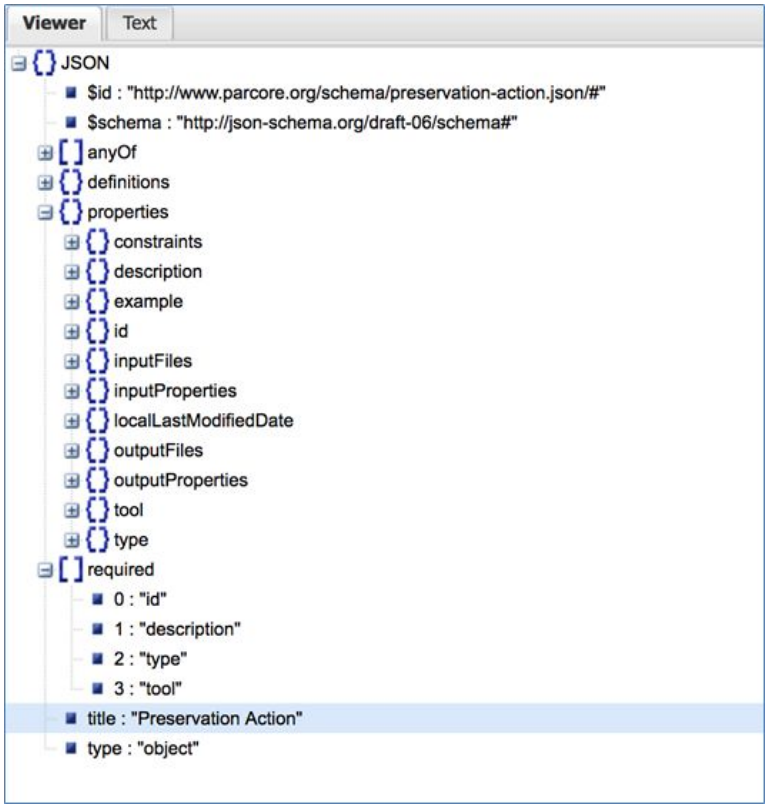
- Property extraction
  - E.g. MediaInfo using EBUCore to get height, width, frame rate
  - Same Tool, same File, multiple Properties, multiple Schema

- Checksum generation and validation
  - E.g. md5sum to generate a MD5 checksum for a file: type and value
  - E.g. md5sum to get PASS/FAIL for given MD5 checksum
  - Same Tool, same File, different Action, Properties go input output

- File format identification
  - E.g. Droid, Siegfried, Fido, File, Tika
  - Business rules on which tools work best and when
  - Business rules on priority/ranking/consensus
  - Is file format a first class entity or just another property?

JSON schemas

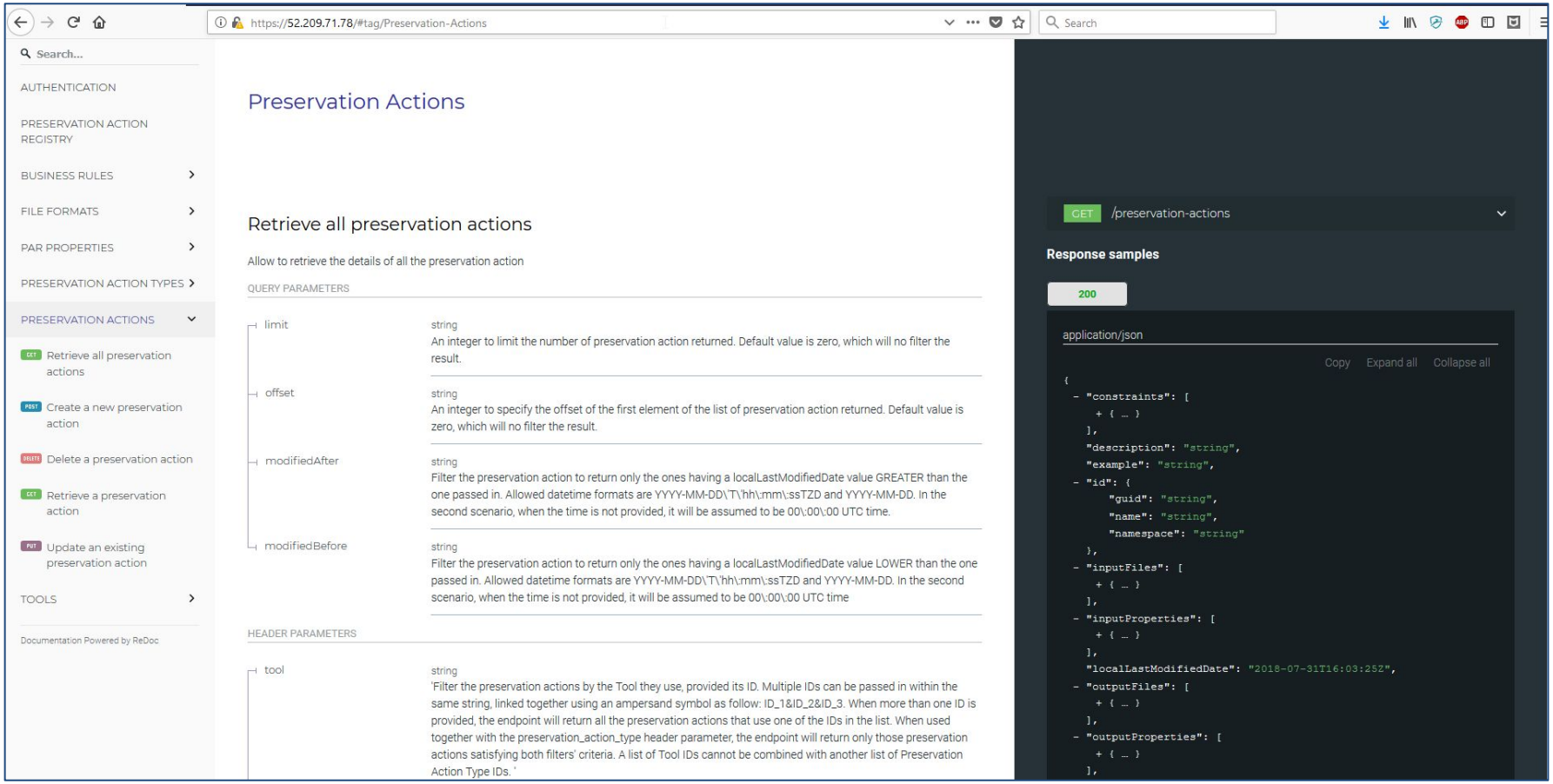
- Tool
- Action
- Action Type
- Format
- Property
- Business Rule



https://github.com/JiscRDSS/rdss-par/tree/master/api

- Lots still to do, e.g.:
  - Objects
  - Containers
  - Linked Data
  - Execution by people as well as machines
  - Avoid combinatorial explosions: actions, properties, tools

APIs



https://github.com/JiscRDSS/rdss-par/tree/master/api

Open Preservation Foundation

arkivum Bringing archived data to life

Preservica Active digital preservation

artefactual®

Jisc

Executable Tool Definitions

- Machine readable spec for running a tool
  - Tool command line
  - Parameters and flags
  - Inputs and outputs
  - Pre and post processing
- Mapping to command lines
  - e.g. extract EBUCore elements/attributes from MediaInfo
  - e.g. extract PRONOM ID from Fido
  - e.g. supply checksum value and filename to md5sum
  - stdin, stdout, stderr, input/output files, parameters and flags, error codes
  - Tools, libraries, dependencies, containers
- Mapping to web services
  - e.g. validators, file format convertors, metadata extractors
  - GET, PUT, POST, sync and async interaction, XML, json, http codes

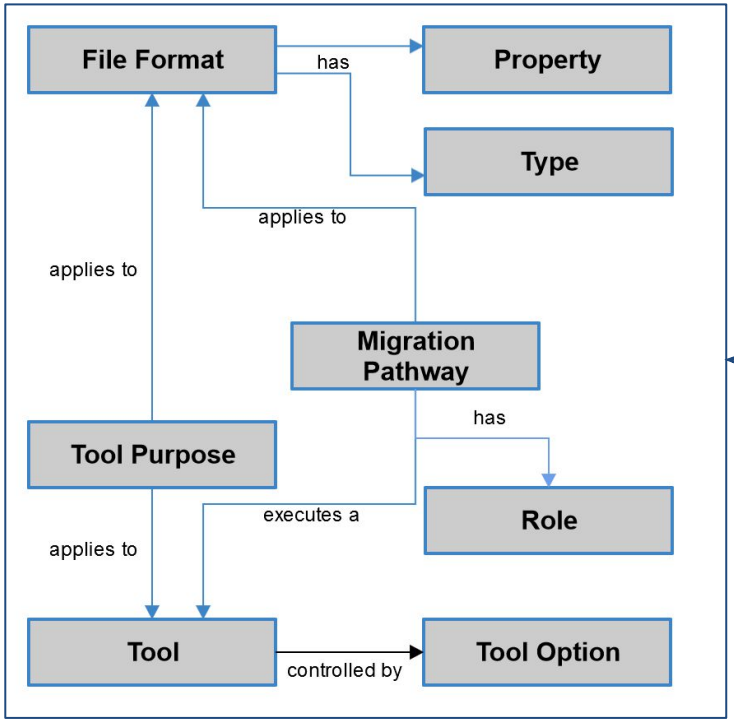
https://github.com/JiscRDSS/rdss-par/tree/master/examples/cwl

Agenda

- Part 1: Background & Motivation
- Part 2: Project outcomes
- Part 3: Proof of concept implementation
- Part 4: Long Term Vision
- Part 5: How can you get involved

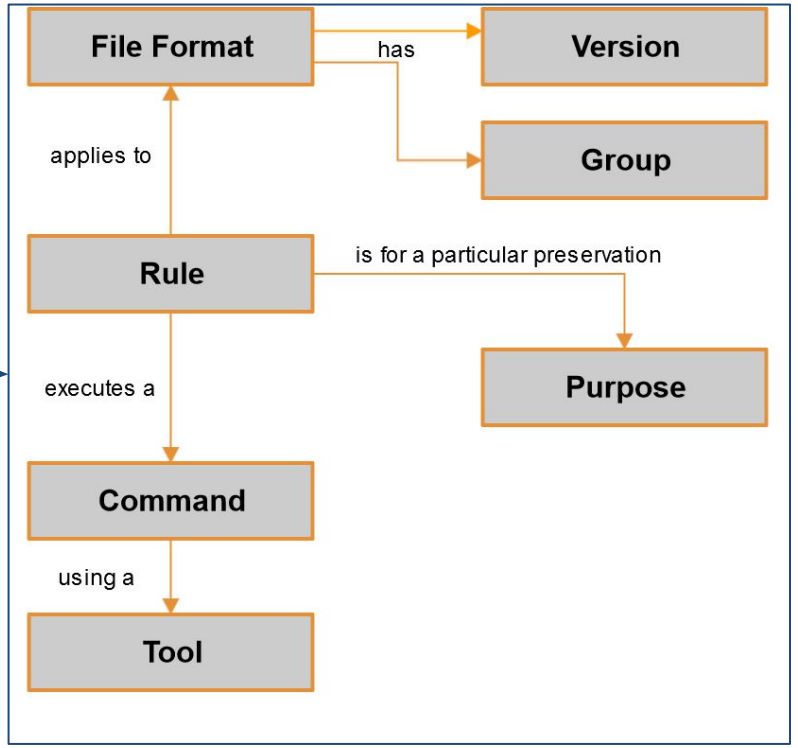
Registry (In)compatibility

Preservica Registry

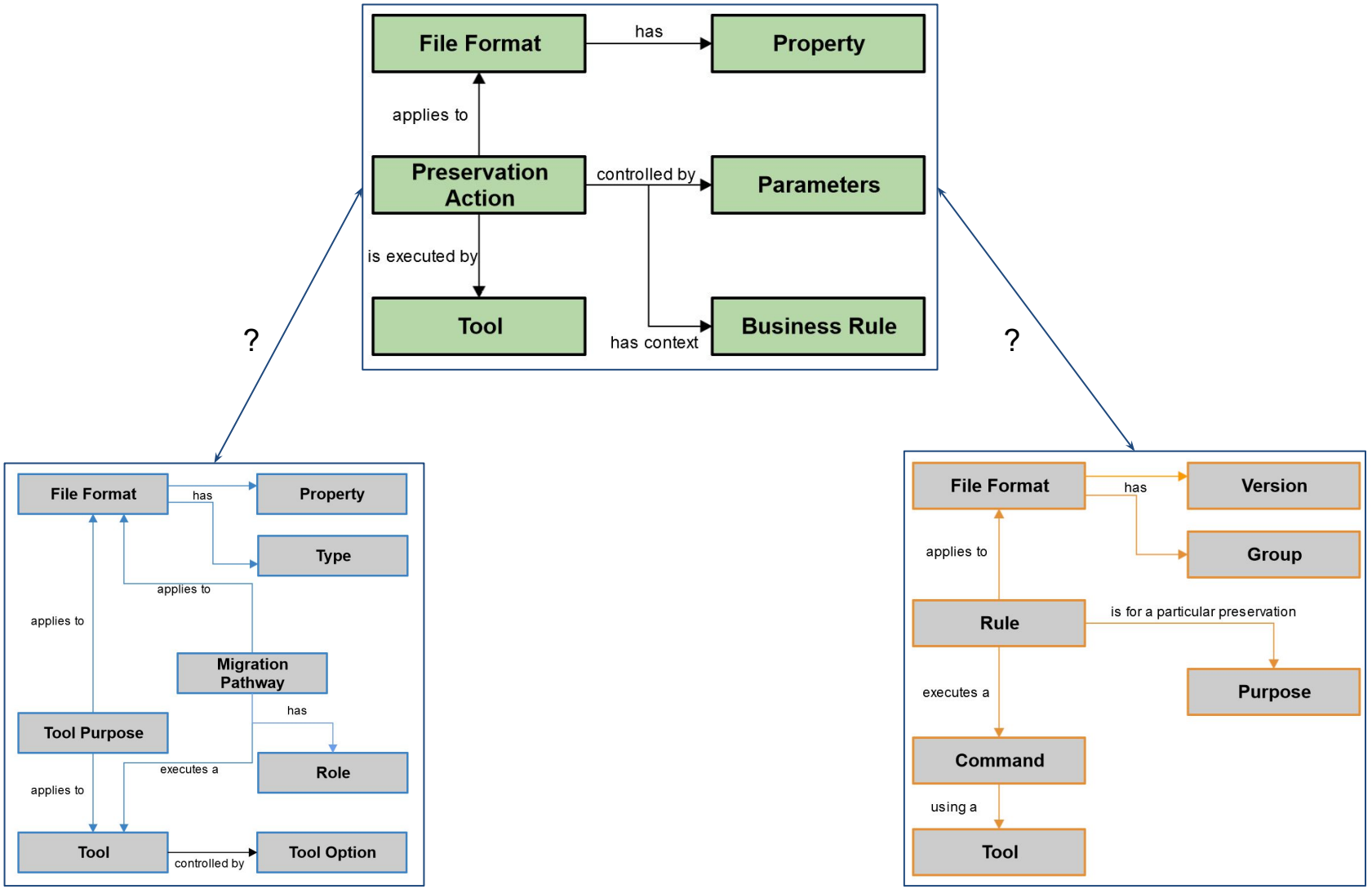


?

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 Archivematica 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 Archivematica's Preservation Planning tab.

Extract Metadata Use Case

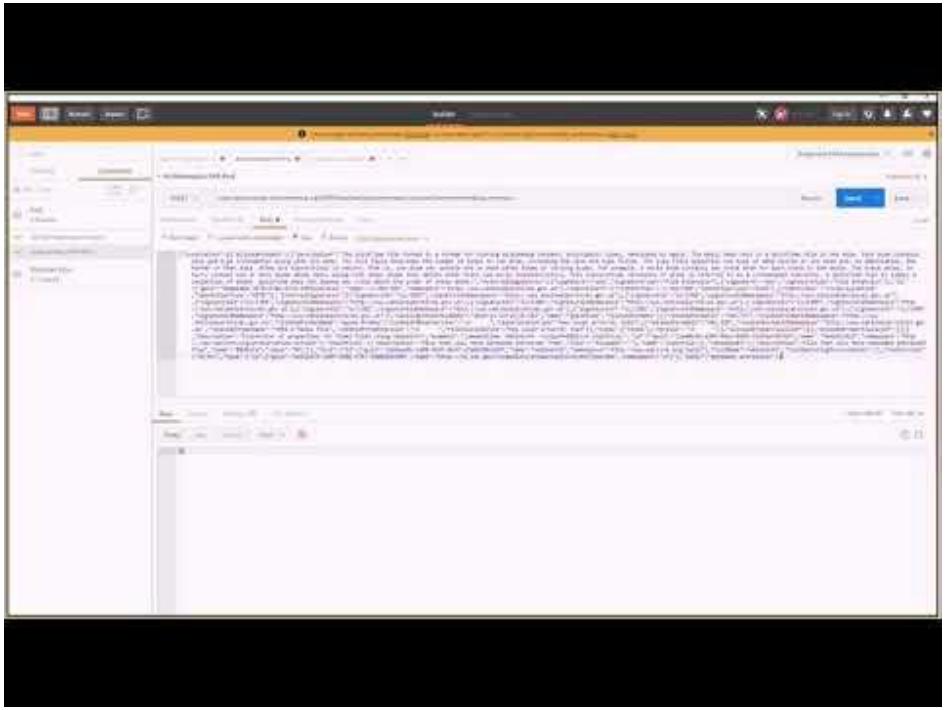
Alice decides that this new capability is something she wants to implement.

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.

Demo



Demo Export and Import of Preservation Action

https://www.youtube.com/watch?v=ELjbgksjBmU

Agenda

- Part 1: Background & Motivation
- Part 2: Project outcomes
- Part 3: Proof of concept implementation
- Part 4: Long Term Vision
- Part 5: How can you get involved

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

Example registry network

Content creation vendor

University

Expert

Research archive

PAR enables all of these links

Government rule

National standards body

Vendor
- Test
- Production

Vendor
- Test
- Production

User
- Manual
- Action

User
- Auto
- Action

User
- Manual
- Action

User
- Override
- Action

User
- Auto
- Action

Expert User
- Evaluate
- Action

Open Preservation Foundation

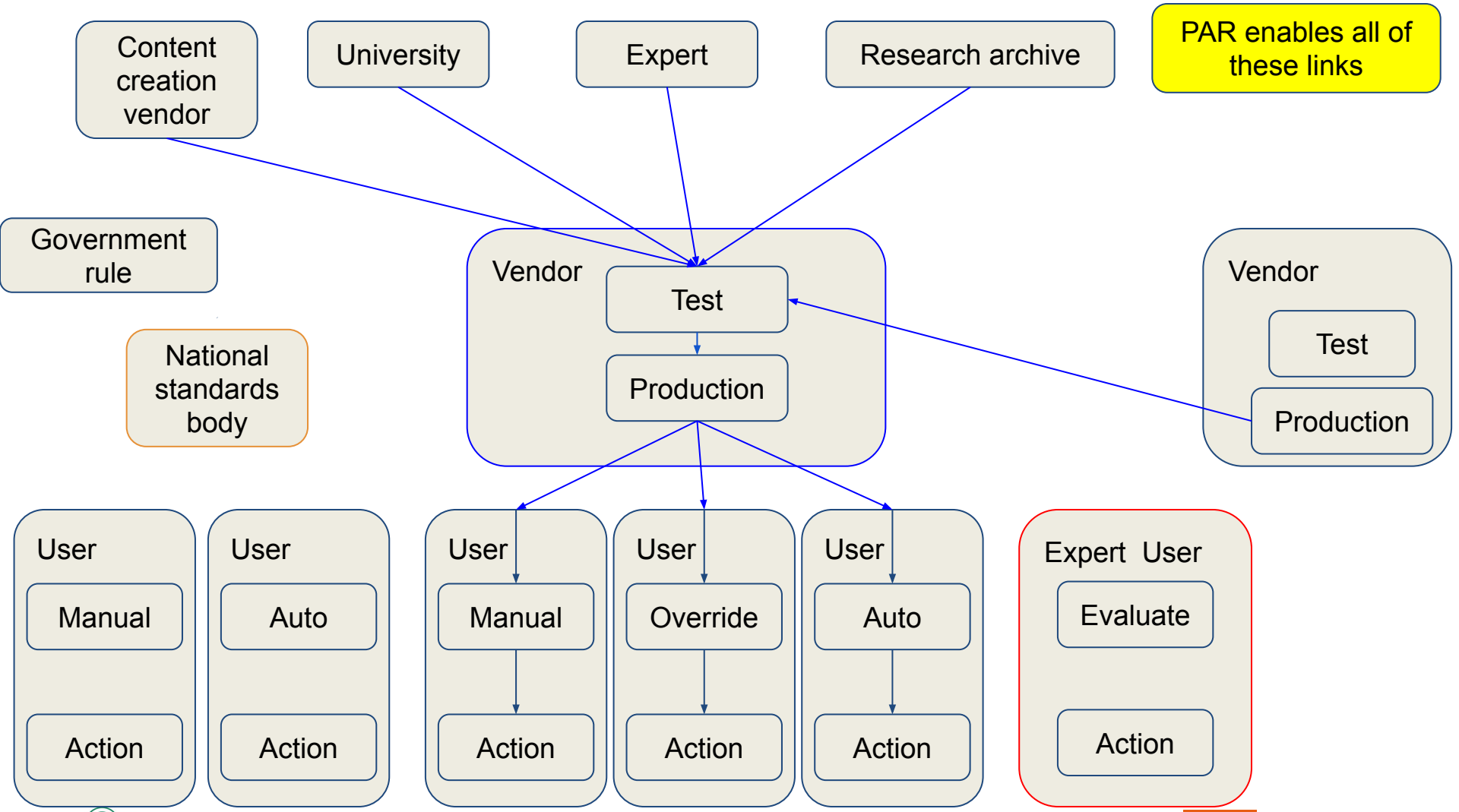
arkivum Bringing archived data to life

Preservica Active digital preservation

artefactual

Jisc

Example registry network



Open Preservation Foundation

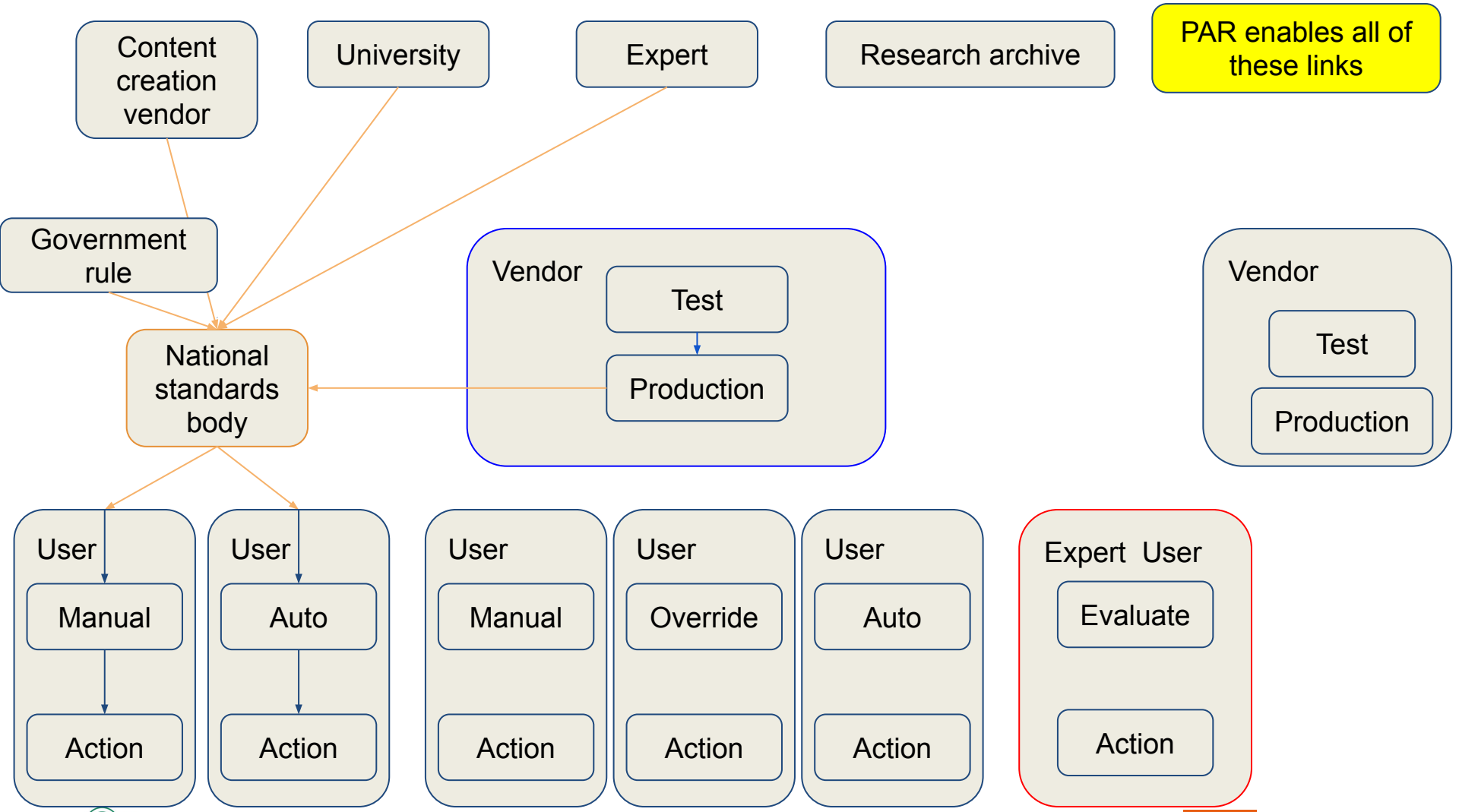
arkivum
Bringing archived data to life

Preservica
Active digital preservation

artefactual

Jisc

Example registry network



Open Preservation Foundation

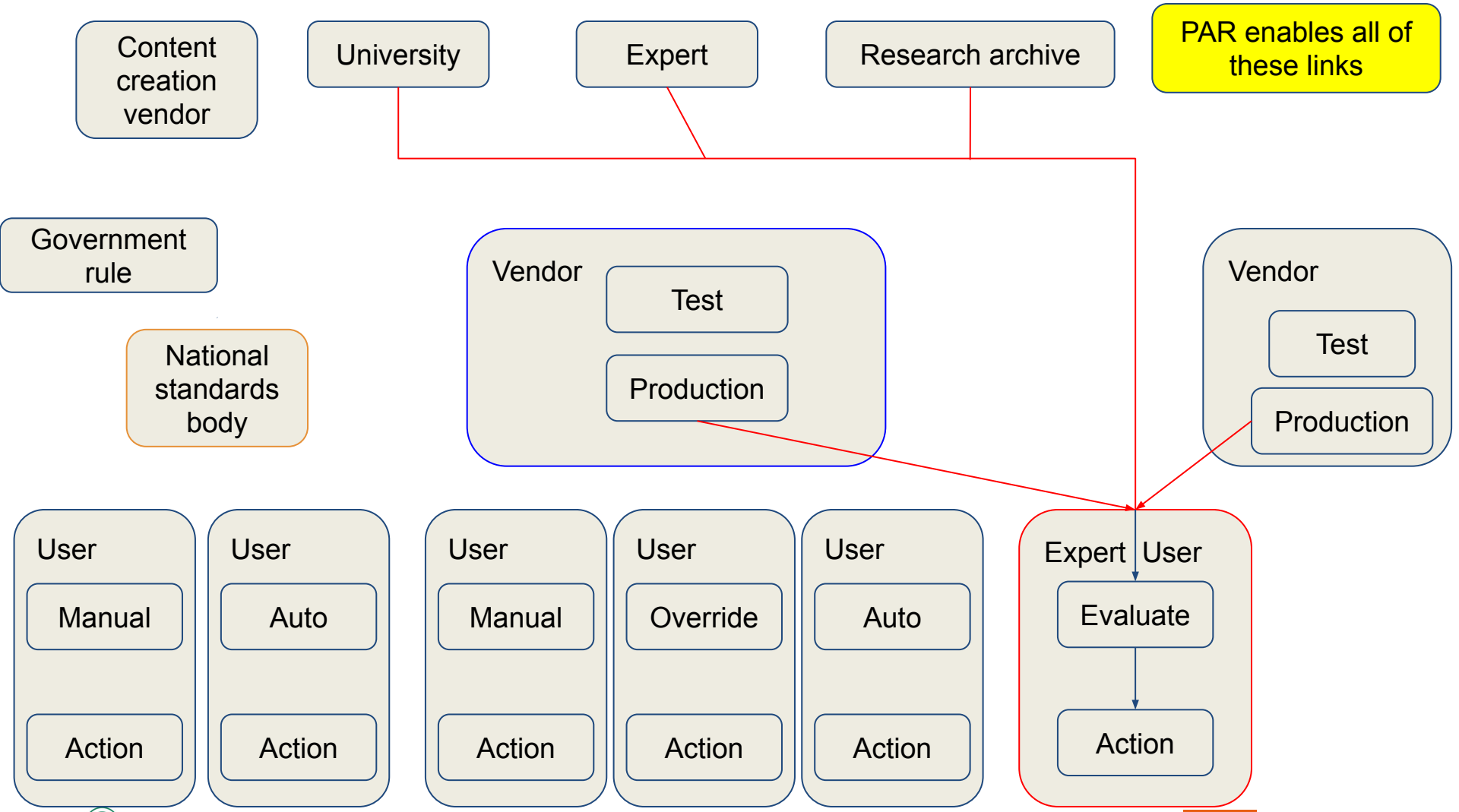
arkivum
Bringing archived data to life

Preservica
Active digital preservation

artefactual

Jisc

Example registry network



Open Preservation Foundation

arkivum
Bringing archived data to life

Preservica
Active digital preservation

artefactual

Jisc

Agenda

- Part 1: Background & Motivation
- Part 2: Project outcomes
- Part 3: Proof of concept implementation
- Part 4: Long Term Vision
- Part 5: How can you get involved

Putting PAR V1 into Production

Technical

- Complete the data model
- Complete the API
- Implement in real systems to do real actions
- Tools to synchronize registries

Community

- Gather user examples
- Identify funders
- Adding endpoints to your registry
- Help code some of the components

Interested?

PAR could enable a global exchange of best practice between all stakeholders in DP

It could define format preservation for the next decade

Further PAR conversation and information to be held in the Library area immediately post this presentation

For further information and to register your interest:

http://bit.ly/2tauvIu

OR

https://openpreservation.org/about/projects/par/

Open Preservation Foundation

arkivum Bringing archived data to life

Preservica Active digital preservation

artefactual

Jisc

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)