

PRESERVATION ACTION RULES WORKSHOP

Parcore: See one, Do one, Teach one

Justin Simpson

Artefactual Systems Inc.
Canada
jsimpson@artefactual.com
<https://orcid.org/0000-0002-1281-1455>

Matthew Addis

Arkivum
United Kingdom
matthew.addis@arkivum.com
<https://orcid.org/0000-0002-3837-2526>

Jack O'Sullivan

Preservica
United Kingdom
jack.osullivan@preservica.com
<https://orcid.org/0000-0002-0306-761X>

Carl Wilson

Open Preservation Foundation
United Kingdom
carl@openpreservation.org
<https://orcid.org/0000-0003-1535-1770>

Sarah Romkey

Artefactual Systems Inc.
Canada
sromkey@artefactual.com
<https://orcid.org/0000-0003-3833-7648>

Jon Tilbury

Preservica
United Kingdom
jon.tilbubry@preservica.com

Abstract – The Preservation Action Registries project (<http://parcore.org>) is developing a common and consistent way to describe and execute preservation policies and actions at a technical level. An initial data model has been created which defines a human and machine-readable way of describing preservation actions and the associated business rules that together make up preservation policies.

Using a skill acquisition technique borrowed from the medical profession (See One, Do One, Teach One) this workshop will lead participants through a three-step process, intended to improve the participants ability to develop and describe their own digital preservation policy in both human and machine-readable forms. The workshop will help participants to better express and share preservation policies in a concise, comprehensive and unambiguous way. Through the participants activities in the workshop, the PAR data model will be tested, validated and further improvements identified.

Participants will be introduced to the PAR project and data model, using examples of working preservation actions and business rules. In smaller group, participants will be led through the process of describing a new preservation policy using the PAR data model. Finally, groups will present their work to the other groups.

Keywords – preservation rule, preservation policy, controlled vocabulary, interoperability, best current practice

Conference Topics – Designing and Delivering Sustainable Digital Preservation; Building Capacity, Capability and Community

I. INTRODUCTION

The Preservation Action Registries (PAR) project was introduced in a paper presented at iPres 2018 [1]. This paper describes a series of problems that arise in the development of digital preservation platforms such as Archivematica and Preservica. These systems lack a common and consistent way to describe and execute preservation policies and actions. Each system must implement some mechanism to define what tools and rules to use when 'doing' digital preservation. There is currently no practical way to share technical information between organisations using different preservation platforms. This limitation presents a barrier to the development of best current practice in the field.

The PAR project proposes a solution in the form of a common data model describing the elements of preservation policies, implemented using machine readable data structures. This data model includes the ability to describe preservation actions that can be performed on digital content, what tools are used to perform these actions, the inputs needed, what results and outcomes are achieved, and in what context the actions should be taken.

Initially funded by Jisc, the project has been taken up by the Open Preservation Foundation. Through a collaboration between three software vendors (Arkivum, Artefactual and Preservica), the project has defined a data model using json-schema [2]. A standardized Application Programming Interface (API) has also been defined [3], using the Swagger API specification language. This allows different digital preservation platforms to implement a common method for sharing this kind of data.

II. WORKSHOP OBJECTIVES

The PAR project has initially focussed on technical interchange of information between preservation systems. The concepts and data model provide a mechanism for organisations to describe digital preservation policies at a detailed and actionable level. In particular, PAR addresses the issue of how to describe the specifics of digital preservation actions so that a description can be taken by someone else in the community and executed either manually or automatically by their preservation system of choice. It is this ‘specification’ aspect of PAR that we will explore in the workshop. The workshop will ask participants to use the PAR concepts and framework to describe their own preservation policies. Through the ‘teach one’ part of the workshop methodology, the descriptions of preservation policies will be ‘tested’ to see if they are sufficiently detailed, have an unambiguous interpretation, and have enough contextual information about when to apply the policies and why. Through this approach, the participants will gain valuable skills described below while the PAR consortium will also be able to answer key questions including:

- Can the PAR model effectively describe real world preservation scenarios?
- What changes are needed to 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?

III. LEARNING OUTCOMES AND BENEFITS

The benefits of the workshop can be split into three areas: benefits to individual participants; benefits to the PAR consortium; and benefits to the wider digital preservation community.

Workshop participants will benefit through:

- Understanding of how to specify preservation policies in a concise, comprehensive and unambiguous way.
- Ability to better communicate and share preservation policies with peers and stakeholders.
- Framework for critical evaluation of their existing preservation policies, e.g. helping to identify gaps, inconsistencies and ambiguities.
- Ability to communicate preservation needs more effectively with vendors or other providers of preservation systems.
- Understanding of how to express the context for applying preservation policies, e.g. as a set of rules and priorities.
- Ability to analyze the technical aspects of preservation in their organization by comparing tools/systems/infrastructure/techniques with good practice of others.

The PAR initiative will benefit through:

- Real-world examples and use cases that will allow us to test the capabilities of the PAR model and identify gaps or problems.
- A set of requirements and priorities for further work.
- Growth of the PAR community and user base.

The digital preservation community will benefit through:

- A published set of preservation policies that cover a range of preservation actions and tools.
- A published set of preservation scenarios of how to describe preservation contexts, e.g. business rules for selecting, comparing and applying policies based on priorities and constraints.
- PAR will become better aligned to and informed by a wide range of real-world preservation use cases.

IV. PRE-REQUISITES

The only pre-requisites for attending the workshop are:

- Participants should bring one or more preservation policies that they already use in their organization (or would like to implement). They should be prepared for these to be described, analyzed and shared in the workshop.
- Participants are expected to actively participate in the workshop, especially in the ‘see one, do one, teach one’ process. For example, all participants will be expected to ‘teach one’ in small groups or one-to-one with other participants.

While PAR supports machine-readable descriptions of preservation actions, e.g. json, we do not expect attendees to take this approach. Instead, attendees will be creating human readable descriptions using the concepts from PAR (actions, tools, rules, formats, objects etc.) and a controlled vocabulary of terms (e.g. preservation action types, file formats, and file properties).

V. WORKSHOP MOTIVATION

Collaborative development of best current practice in any field is an ambitious and difficult undertaking. A common method for describing and implementing digital preservation techniques should reduce the cost of developing best current practice, and reduce the learning curve for individuals and organizations attempting to improve their knowledge and expertise in the field.

The PAR project offers a starting point for developing a common method or approach. The project needs input from a wider audience of digital preservation practitioners to validate the approach, test and improve the existing data model and api, and to develop a corpus of preservation

actions and business rules that represent actual best current practice in the field.

Practitioners will need to learn about the PAR data model and develop expertise in working with it. We propose to borrow a skill acquisition technique that is well known in the medical profession, commonly referred to as ‘See One, Do One, Teach One’ [4]. We believe this simple methodology allows the greatest chance for the concepts and techniques of the PAR project to be shared with a new audience.

VI. WORKSHOP STRUCTURE

A. *Introductions and overview of PAR*

The PAR consortium will provide an overview of the PAR project, its objectives and results so far. This will include a review of the PAR data model and how this can be used to define preservation policies based on preservation actions, tools and rules. The PAR data model will provide the framework for the See one, do one, teach one part of the workshop and allow project participants to describe their preservation policies in a consistent way.

B. *See one*

The PAR consortium will provide the attendees with ready-worked examples of how the PAR approach can be used to describe preservation policies. These will be based on user scenarios (e.g. ‘An organization is responsible for preserving audiovisual content and wishes to use EBUCore to standardize the way it describes its assets’) and the specific preservation actions needed to achieve this (e.g. Use the MedialInfo tool to characterize the technical aspects of video files and store the results as EBUCore in XML format). Examples will cover several of the main PREMIS preservation event types (e.g. fixity, characterization, validation, format normalization). This will enable the attendees to ‘see one’ before they are asked to develop their own descriptions.

C. *Do one*

Attendees will be split into small groups and asked to select one or two preservation policies to work on from the candidates brought by the attendees. We will group attendees with similar policies. Each group will discuss the policies and develop a description using the PAR framework. A member of the PAR consortium will facilitate each group. Each group member will be involved in the ‘Do one’ activity, e.g. by focusing on different aspects of the description such as the business rules, tool specification, action definition etc. GoogleDocs or similar will be used for collaborative authoring and to help the PAR consortium to capture the results of the group work.

D. *Teach one*

Each group member will present their group’s description of a preservation policy to a member of another group (teach one). This will be done in pairs or in small groups. Everyone will get the chance to ‘teach’ as well as listen and comment on other people’s policy descriptions. Comments will be captured in shared GoogleDocs alongside

the preservation policy descriptions. We plan to identify where descriptions are incomplete, ambiguous or can be better specified. This will test whether PAR can capture all information necessary to enact a preservation action in a real-world environment. This part of the PAR workshop will focus on the ‘teaching’ aspect of ‘see one, do one, teach one’ and the use of question/response as part of teaching as a way to validate that the information being exchanged between the ‘teacher’ and the ‘student’ is clear, concise and complete.

E. *Feedback and next steps.*

The final stage of the workshop will be used to capture general comments and feedback on the participants experience with the PAR approach. This will be done as a plenary involving all participants so everyone has chance to contribute and share their suggestions for PAR improvements.

REFERENCES

- [1] Digital Preservation Interoperability through Preservation Action Registries https://figshare.com/articles/Digital_Preservation_Interoperability_through_Preservation_Actions_Registries/6628418
- [2] PAR Json Schemas, Github <https://github.com/JiscRDSS/rdss-par/tree/master/schemas>
- [3] PAR REST API Specification, Github <https://github.com/JiscRDSS/rdss-par/tree/master/api>
- [4] See One, Do One, Teach One – Not Just for the Medical Profession <https://spin.atomicobject.com/2011/04/29/see-one-do-one-teach-one-not-just-for-the-medical-profession/>