Digital Preservation as a dependency

DPC Members Forum

David Cirella, Yale Library

2025-04-25

Introduction

1. Digital Preservation at Yale
2. digital preservation as a dependency
3. (An) AV Life-cycle
4. Supporting dependents
    - Software development
    - Support for developers
    - Support for users
5. Results
6. Reflections

Background / timeline

- **2013** – Digital Preservation program starts
- **2013 - 2015** – Infrastructure planning and development
- **2015 - 2016** – System selection
- **2016** – Preservica installation
- **2018** – Team expands +2 Digital Preservation Librarians
- **2020 - 2021** – Demand for Preservica grew during lockdown
- **2023** – Team expands +2 (Born Digital Specialist & Digital Preservation Librarian)
- **2024** – Dept. reorg, new Head of Digital Preservation

Digital Preservation at Yale

Team of 5 serving

- Libraries and Archives
- Law Library
- Museums and Galleries
- University Archives

Services

- Digital preservation system
  - administration and maintenance
  - training
  - user support
  - bulk packaging, ingest, & access
  - integration support
- Born digital accessioning and access

Digital Preservation System at Yale

- Preservica EE on-prem
- Ingest workflows: > 1.1 million
- Current holdings:
  - > 138 million files
  - > 1.20 PiB (unique)
  - With redundant copies:
    - > 275 million files
    - > 2.47 PiB

- Infrastructure Support:
  - LibraryIT - DevOps
  - Central IT
- Storage infrastructure:
  - On-prem ssd/hdd
  - On-prem Tape
  - Hybrid on-prem cache/cloud
  - Cloud

```
digital preservation == dependency
```

```
digital preservation != obstacle
```

(An) AV Life-cycle

Action:

1. Physical item selected
2. Digitization
3. Packaging
4. Ingest
5. Access

Stakeholder:

`Collecting Unit`

`Media Preservation`

`Digital Preservation`

`Media / Digital  Preservation`

`Digital Special Collections Access`

(An) AV Life-cycle w/ integrations

Action:	Stakeholder:	System:
1. Physical item selected	`Collecting Unit`	-
2. Digitization	`Media Preservation`	ICE (*digitization wf system*)
3. Packaging	`Digital Preservation`	AVPackager
4. Ingest	`Media / Digital Preservation`	ICE (*digitization wf system*)
5. Access	`Digital Special Collections Access`	Preservica
		Aviary (*AV Access platform*)

digital preservation != obstacle

Software integration development

- Collaborators wrt stakeholder specifications/requirements

Operations and deployment planning

Consultative role with external developers

- Collaborators wrt stakeholder processes/practices

End user support

- Collaborators wrt stakeholder use-cases

DP's role - AV Life-cycle

Receive product of digitization

Create specified package

Facilitate ingest management & tracking by external system

Maintain interface to data and linking to descriptive systems

Software integration development

AVPackager v1

- **Input:** product of AV digitization
- **Output:** multi-representation SIP
- Specified by Yale, developed by Preservica
    - Leveraged dp system interfaces
- Package and ingest, one action
- Original implementation relied on (now) deprecated functionality

> **Collaborators:** DP system vendor, Media Preservation, Special Collections

Software integration development

AVPackager Re-Development

- Drop-in replacement
- Specified/designed by Yale DP, developed by Library SW Engineering
- Allowed growth in functionality
- Ability to rapidly develop and test
- Reduced complexity

> **Collaborators:** Library SW Engineering, Media Preservation, Special Collections

Software integration development

AVPackager

Input:

```
└── yul_avpackager_12_52b15044
    ├── yul_avpackager_12_BACK.jpg
    ├── yul_avpackager_12.dm.mov
    ├── yul_avpackager_12.dm.mov.sha512
    ├── yul_avpackager_12.jpg
    ├── yul_avpackager_12.mp4
    ├── yul_avpackager_12.mp4.sha512
    ├── yul_avpackager_12.mxf
    ├── yul_avpackager_12.mxf.sha512
    ├── yul_avpackager_12.pdf
    └── yul_avpackager_12.xml
```

Output:

```
├── bea9797c-2601-11ec-bea5-d3631a6dd69e
│   ├── content
│   │   ├── yul_avpackager_12_BACK.jpg
│   │   ├── yul_avpackager_12.dm.mov
│   │   ├── yul_avpackager_12.jpg
│   │   ├── yul_avpackager_12.mp4
│   │   ├── yul_avpackager_12.mxf
│   │   └── yul_avpackager_12.xml
│   └── metadata.xml
└── bea9797c-2601-11ec-bea5-d3631a6dd69e.protocol
```

Software integration development

AVPackager - Interface

`.properties`

```
du.catalogueref=AVTest1
du.scope=AVTest1
du.title=AVTest1
security.tag=OPEN
aspace.id=archival_object_000000
source.location=/opt/import/sample_data/AVTest1
```

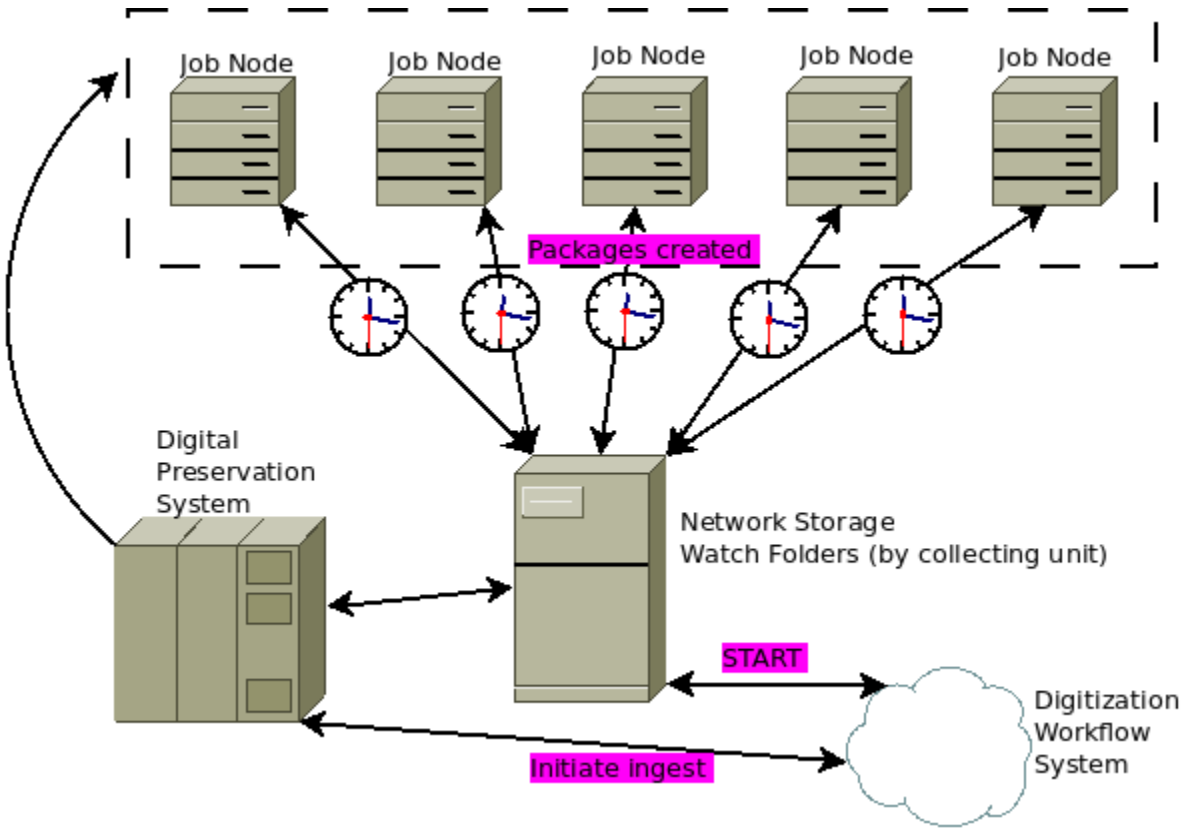
Operations and deployment planning

AVPackager Re-Development

- Identify scheduling method
- Collaborate on deployment strategy
- Acceptance and regression testing

> **Collaborators:** Library DevOps, Library SW Engineering

Packaging and Ingest



Developer Consultation

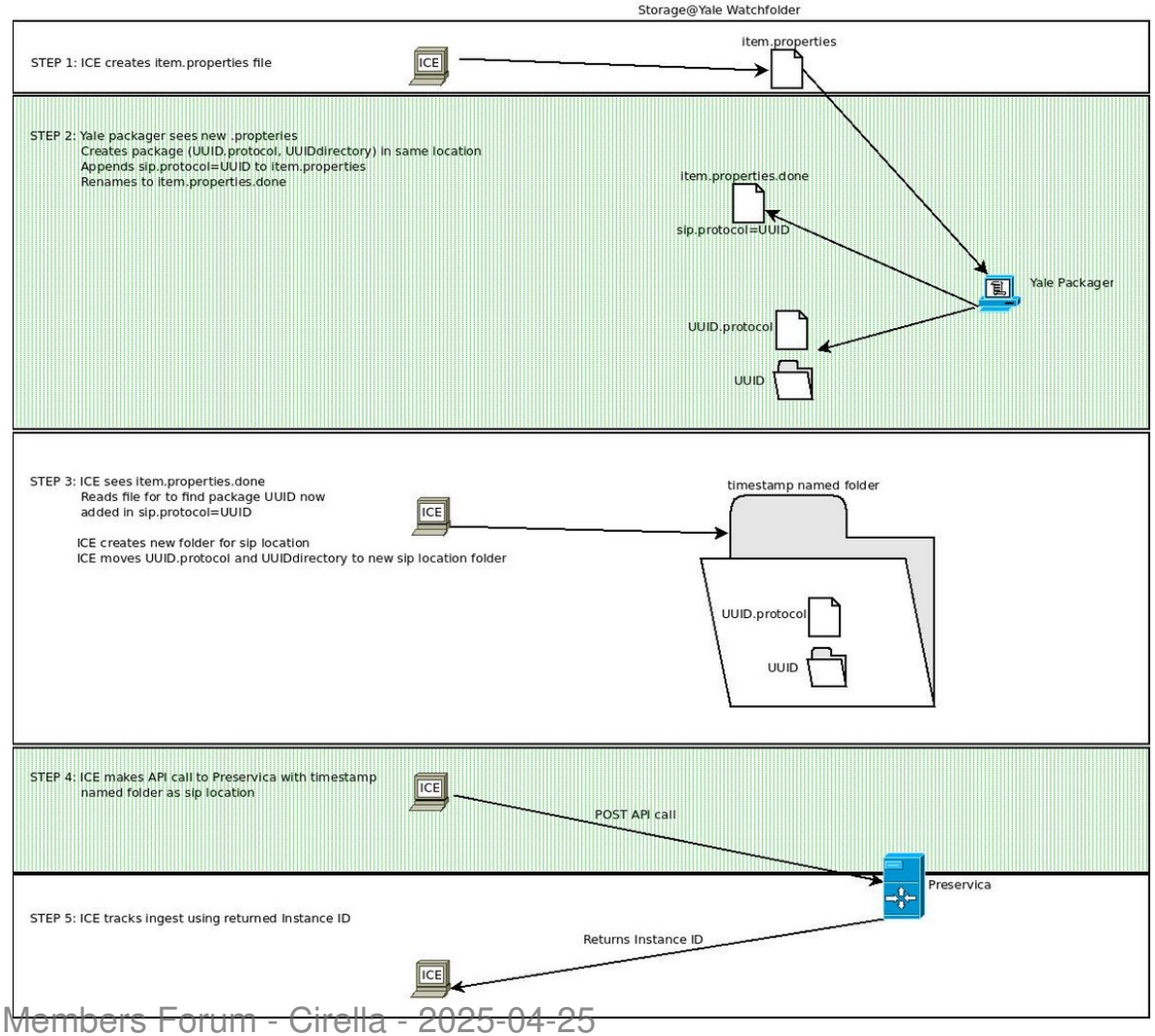
Digitization workflow system

- Provide data model testing against requirements
- Identify/propose API interactions
- Testing and monitoring through development cycle

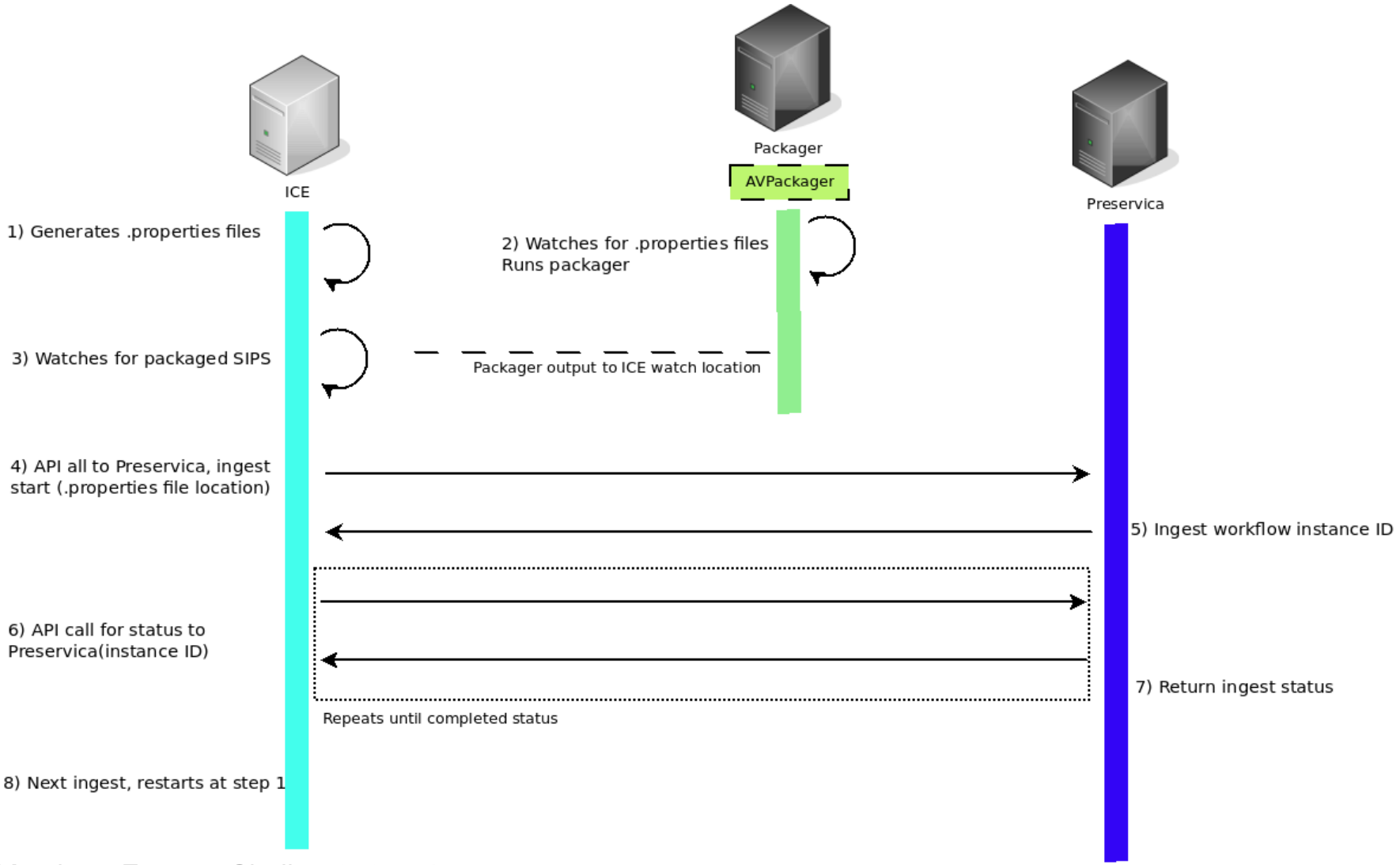
> **Collaborators:** Workflow system vendor, Media Preservation

Developer Consultation

ICE, Yale Packager, Preservica v6 interactions -



Developer Consultation



Developer Consultation

Access system

- Open decisions around data modeling
- Consult with Access system vendor
  - Provide system interaction overviews
  - Example API calls, data model traversal
- Arrange for dev/test system access
- Create test data sets, complete testing
- Open/ongoing analysis/troubleshooting

> **Collaborators:** Access system vendor, Special Collections

End/staff user support

- Monitor/report on ingests
- Use case analysis/consultation
- Front-line support for users
- Training
- Analysis of related system interaction

> **Collaborators:** Media Preservation, Special Collections Access, Collecting Units

Results

Digital preservation as a dependency

- 24/7 packaging/ingest of digitized AV
  - > 135 TB , > 93,000 files
- Roughly 635 TB, ~30,000 (multi-file) packages structured for Access ingest
- Active AVPackager development
  - Added features
  - Bug fixes
  - Enhanced logging

Reflections

- Model of open operation/collaboration/support
- High profile use of digital preservation services
- Commitment to 'source of truth'
    - Long-term sustainability/efficiency
- Development of deep technical understanding
    - Advocate for digital preservation best practices
    - Validate at-scale
- Replicate-able model
    - Similar workflow for Digital Collections System

Questions

Thank you!