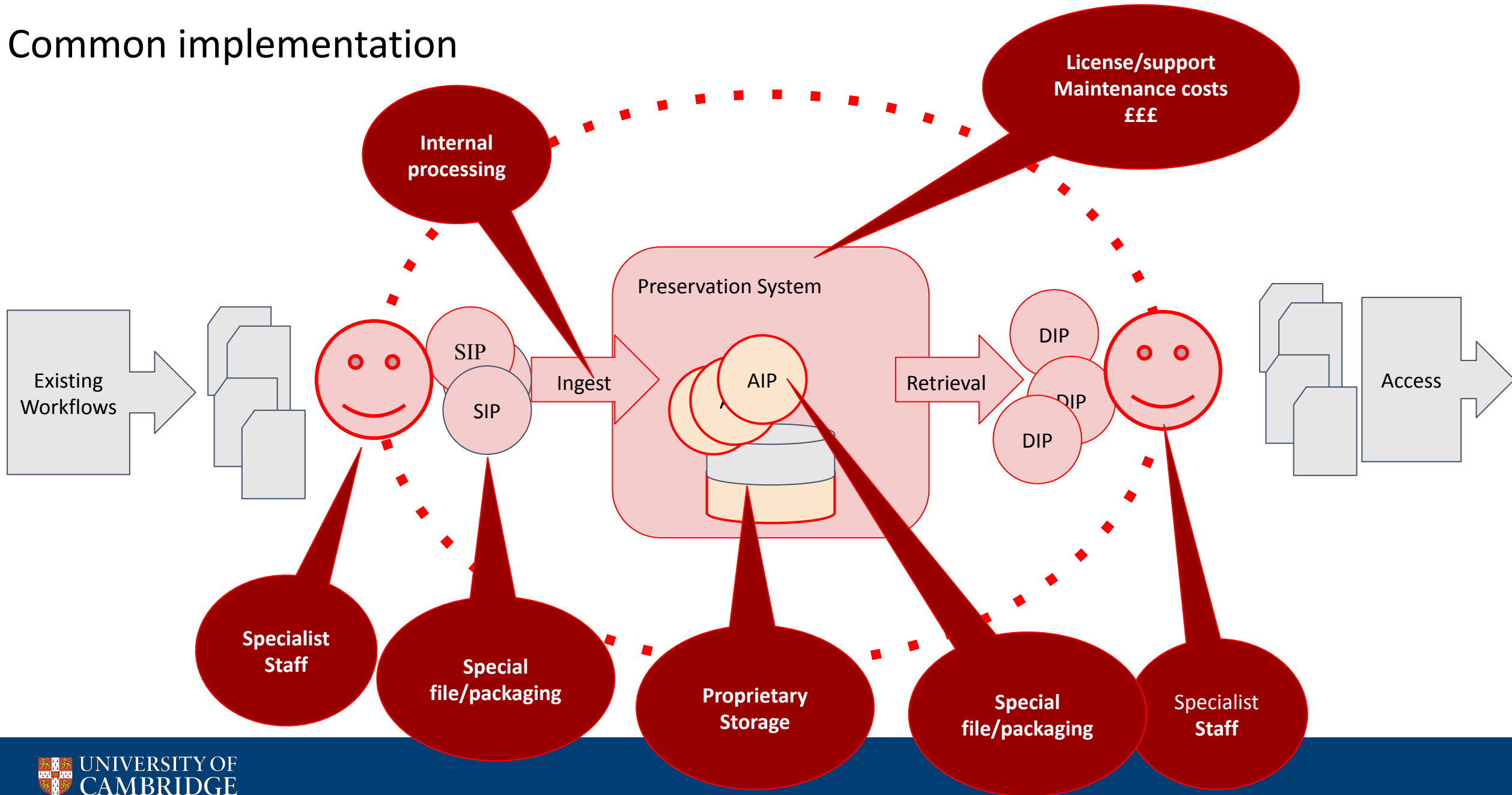


Sustainable Digital Preservation

John Gostick, Technical Lead
Tuan Pham, Technical Director and Senior Responsible Supplier

Digital Preservation Programme,
Cambridge University Library

Common implementation



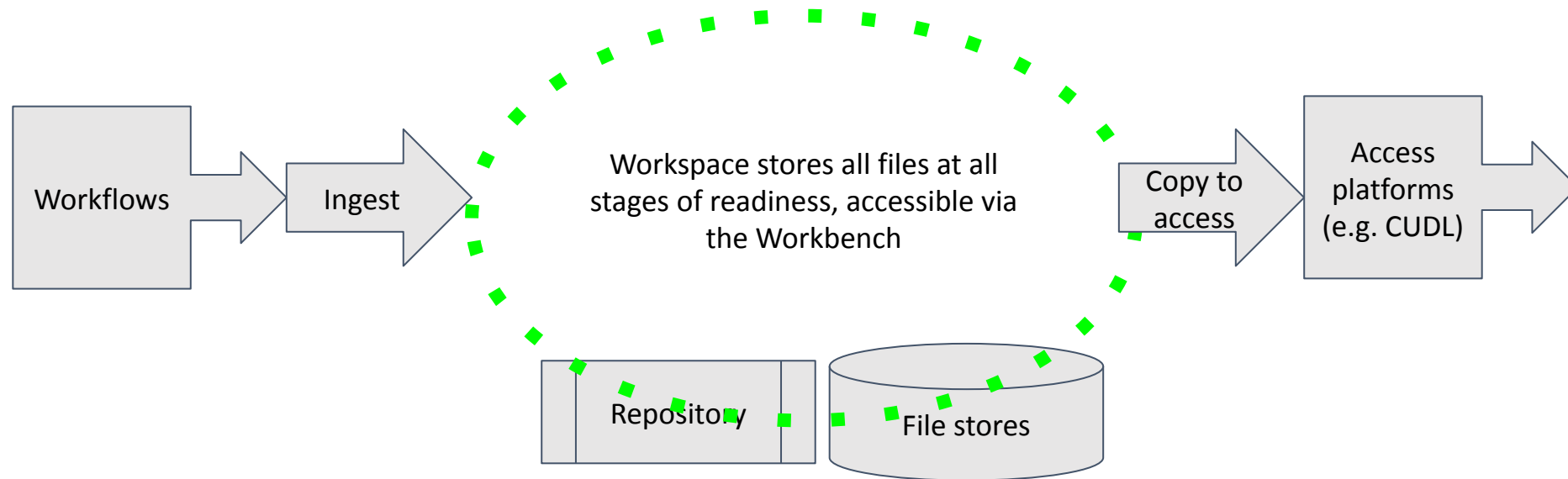
Hard to predict availability of resources in long term



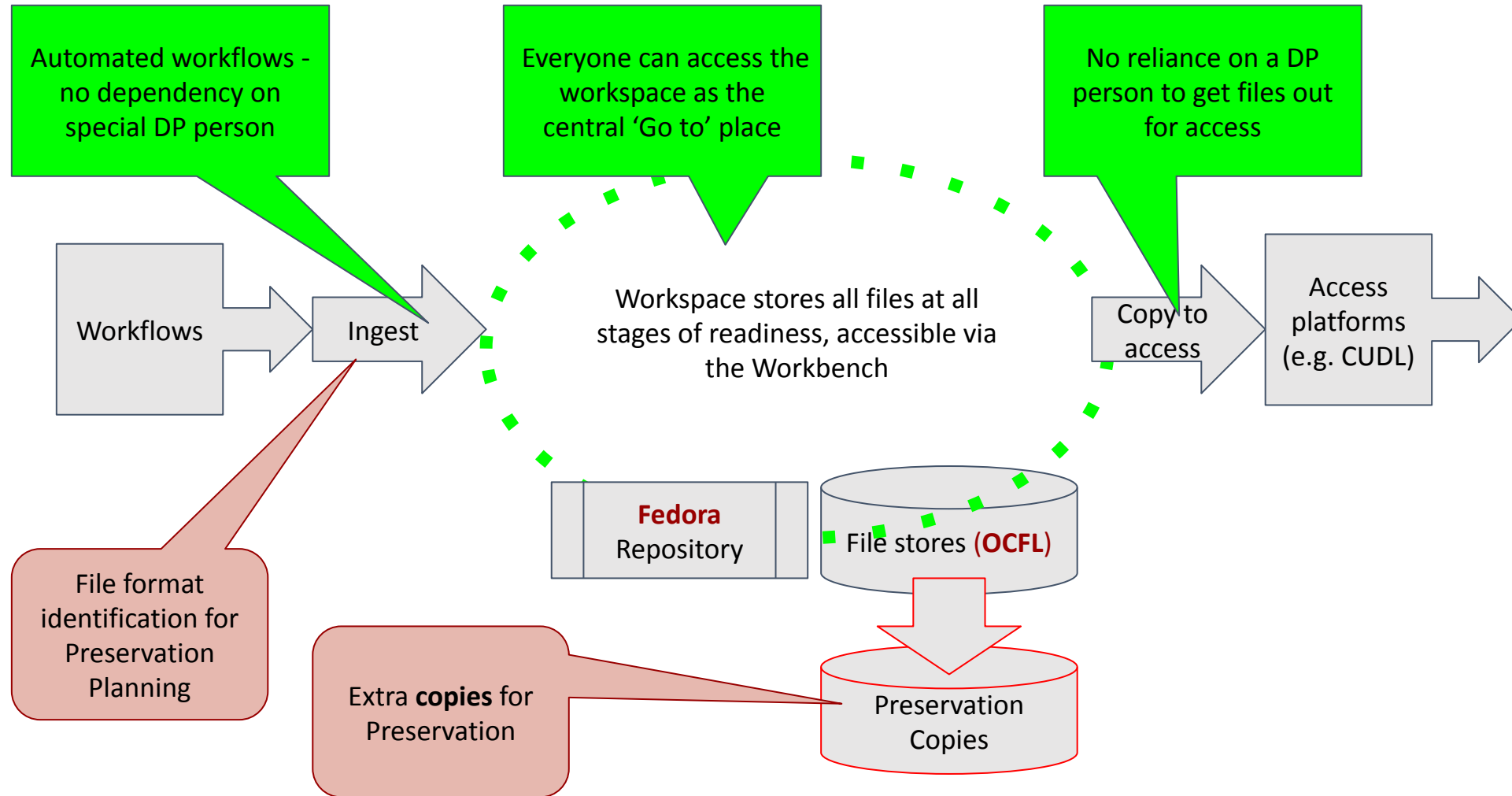
- Cost of licenses/maintenance - what happens if budget is limited? Exit?
- Will specialist DP staff be around?
- Deal with requirement changes? Need developers with special skills?
- Scale up? And down?

Sustainable => less dependent on variables that cannot be controlled / predicted

Collection management



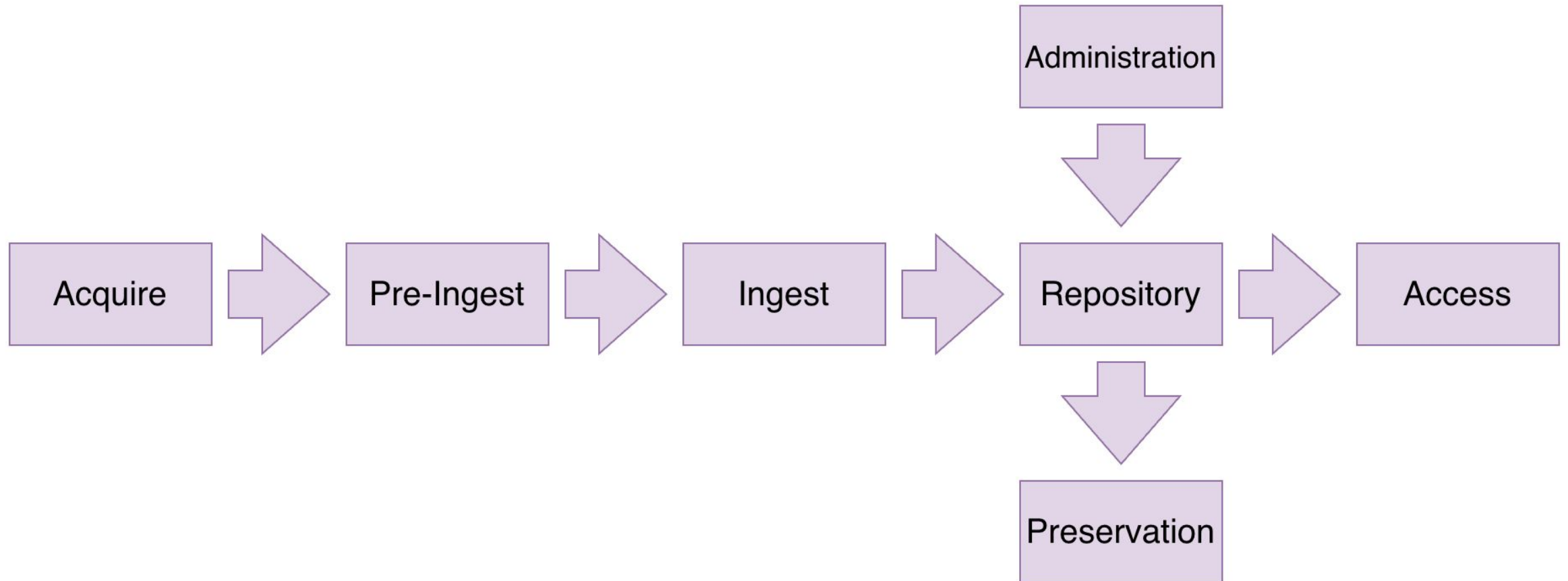
Collection management **with Preservation**



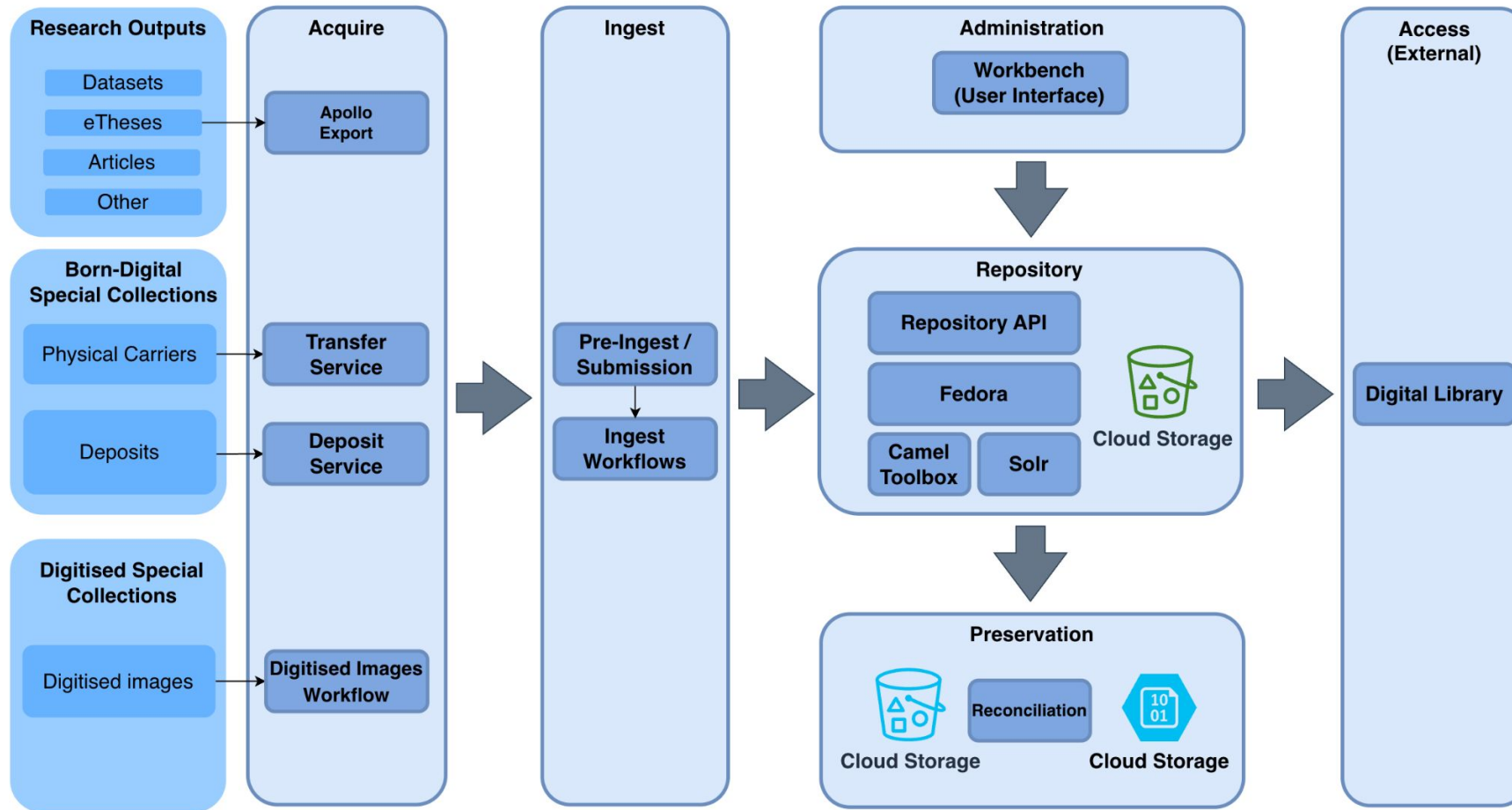
Cambridge's approach

- Embed preservation activities into normal workflows. No specialist knowledge of DP required.
- Open environment - repository is a workspace, open and accessible to all via the 'Workbench'. Inclusive - everyone have same knowledge.
- Open-source software, internalised and adapted.
- Open standards, using OCFL as storage, all files / data human-readable.
- Built with microservices - adaptable, flexible to allow for future changes without significant impact.
- Scale services up as required, down to minimum when idle.

DPS Workflow



DPS Components

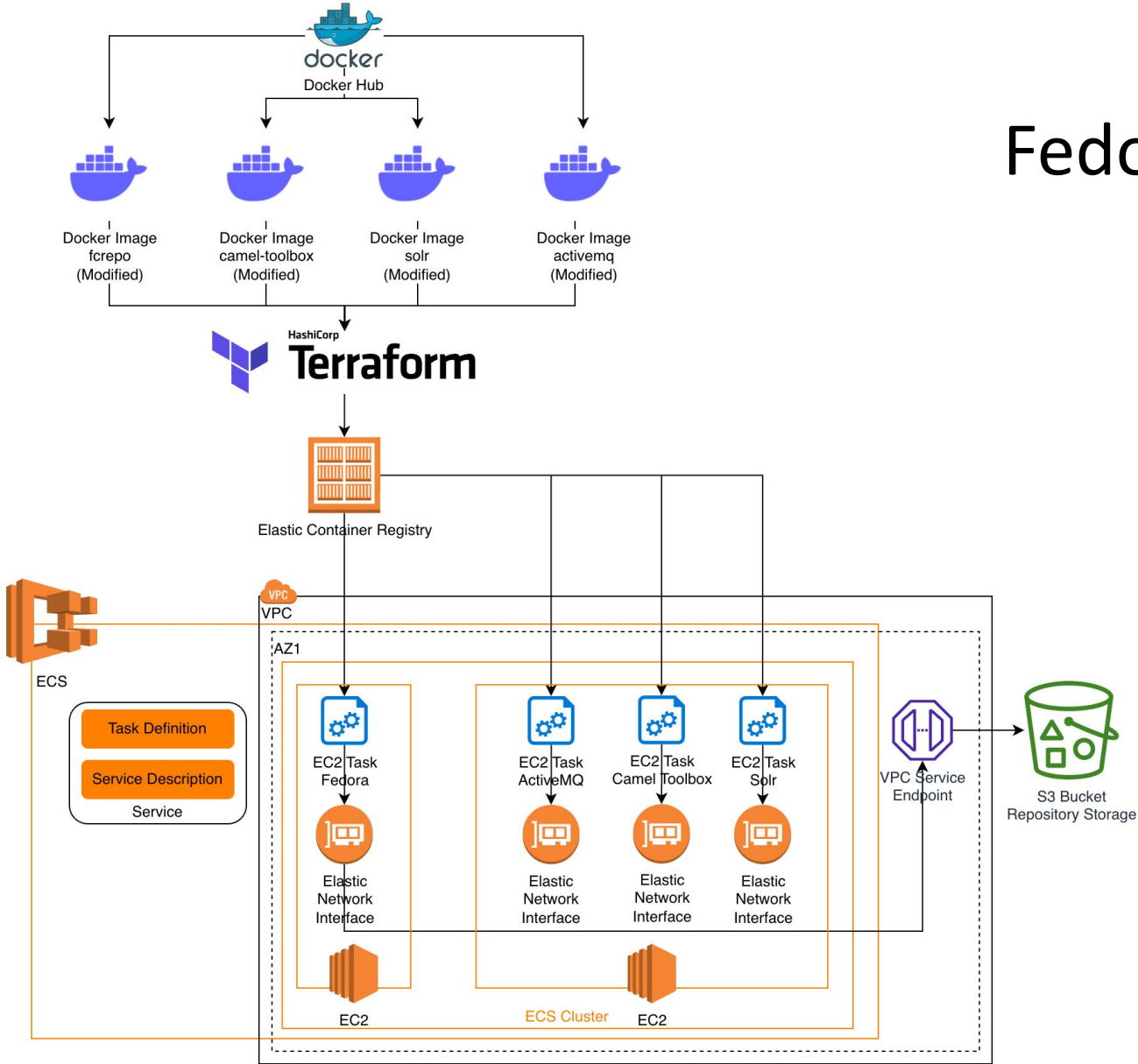


Repository: Fedora 6

- standards-based and open-source
- OCFL on s3
- supports versioning and mutable head extension
- extremely flexible, supports any metadata schema
- stores metadata in OCFL alongside binary data
- supports transactions for atomic operations
- verifies checksums on submission



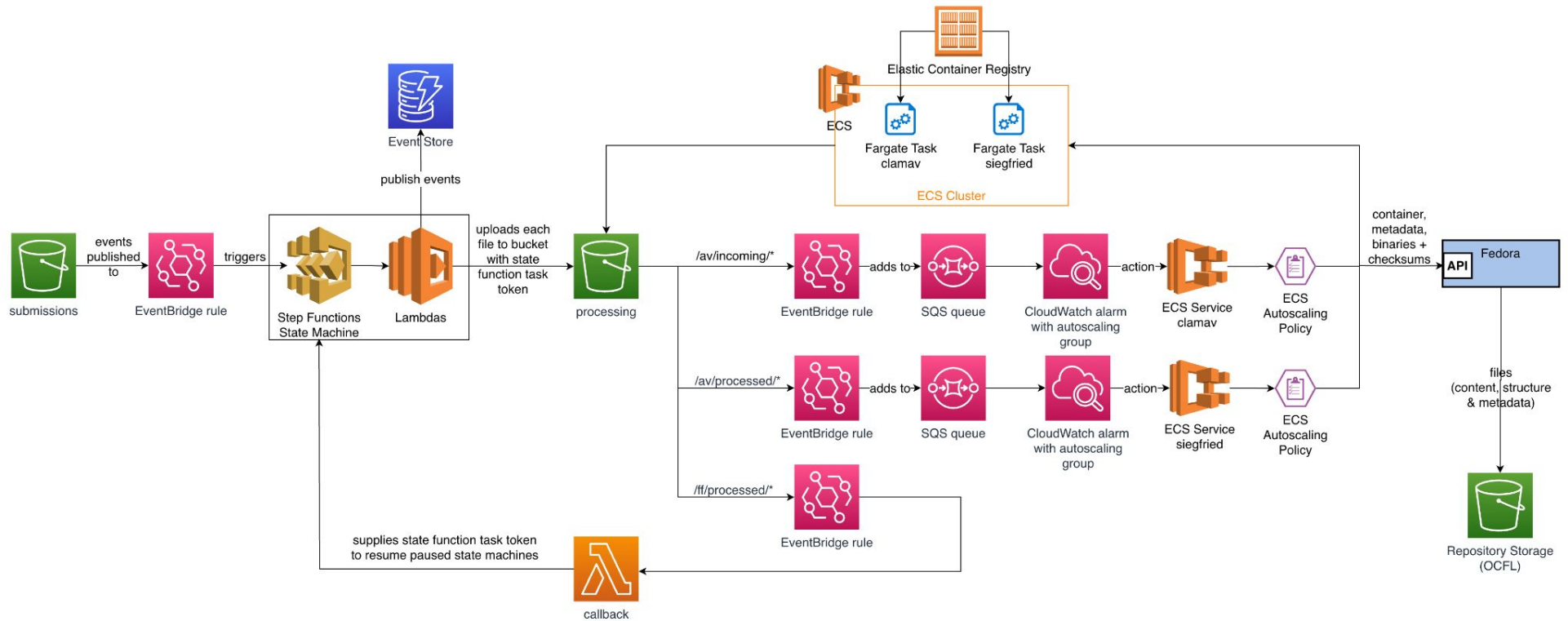
Fedora: AWS Architecture



Ingest Workflow: What does it do?

- Check for duplicate submissions
- Unpack submission package and embedded containers
- Extract any supplied metadata for package
- Scan for viruses
- Identify file formats
- Extract embedded metadata and language from files
- Verify then store files and metadata in repository (Fedora)
- Uses Fedora transactions to ensure completeness
- Create a version in Fedora/OCFL
- Update workflow database table
- Modular, shared components support multiple workflows
- Allows ingest for appraisal and accessioning before preservation

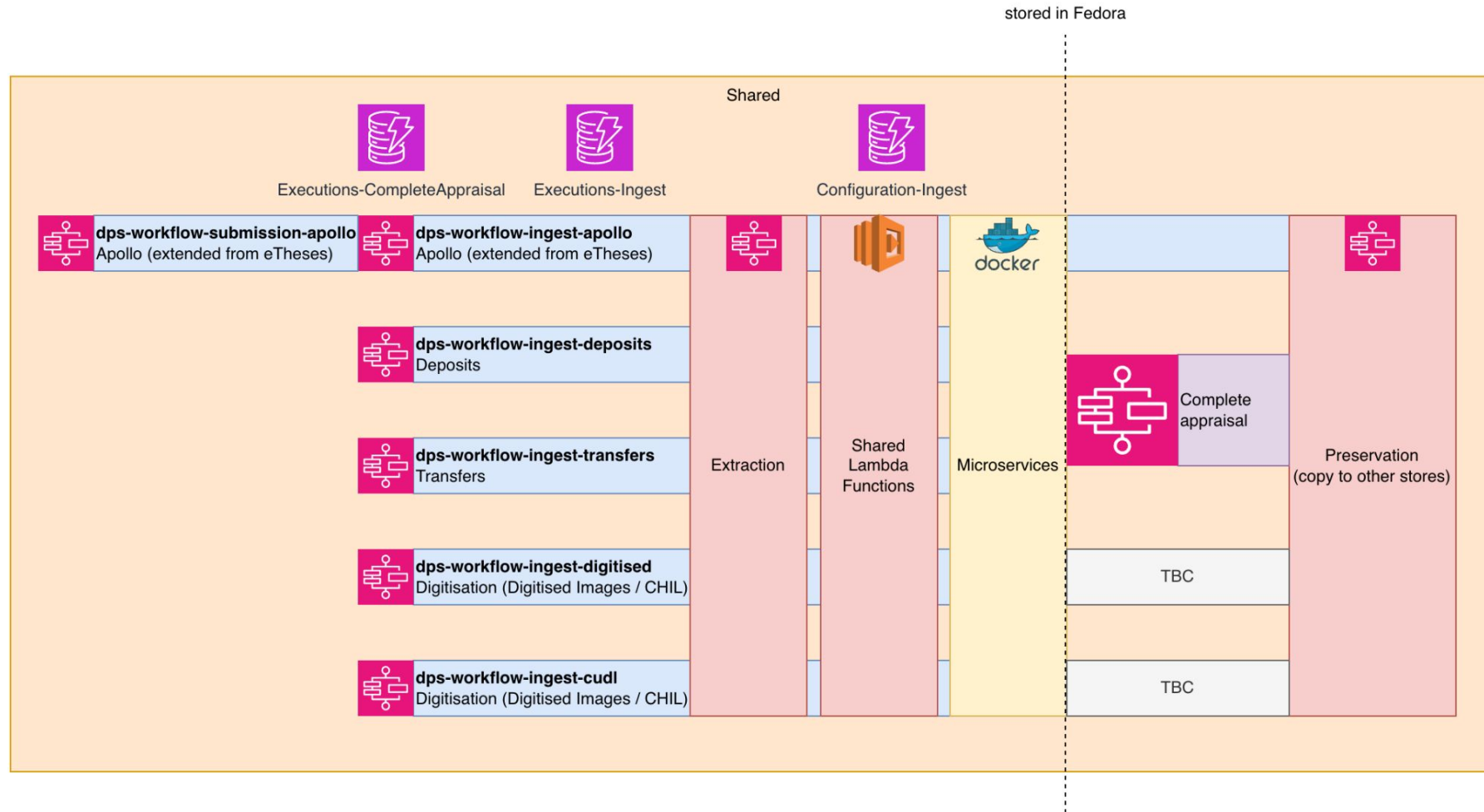
Ingest Workflow: AWS Architecture



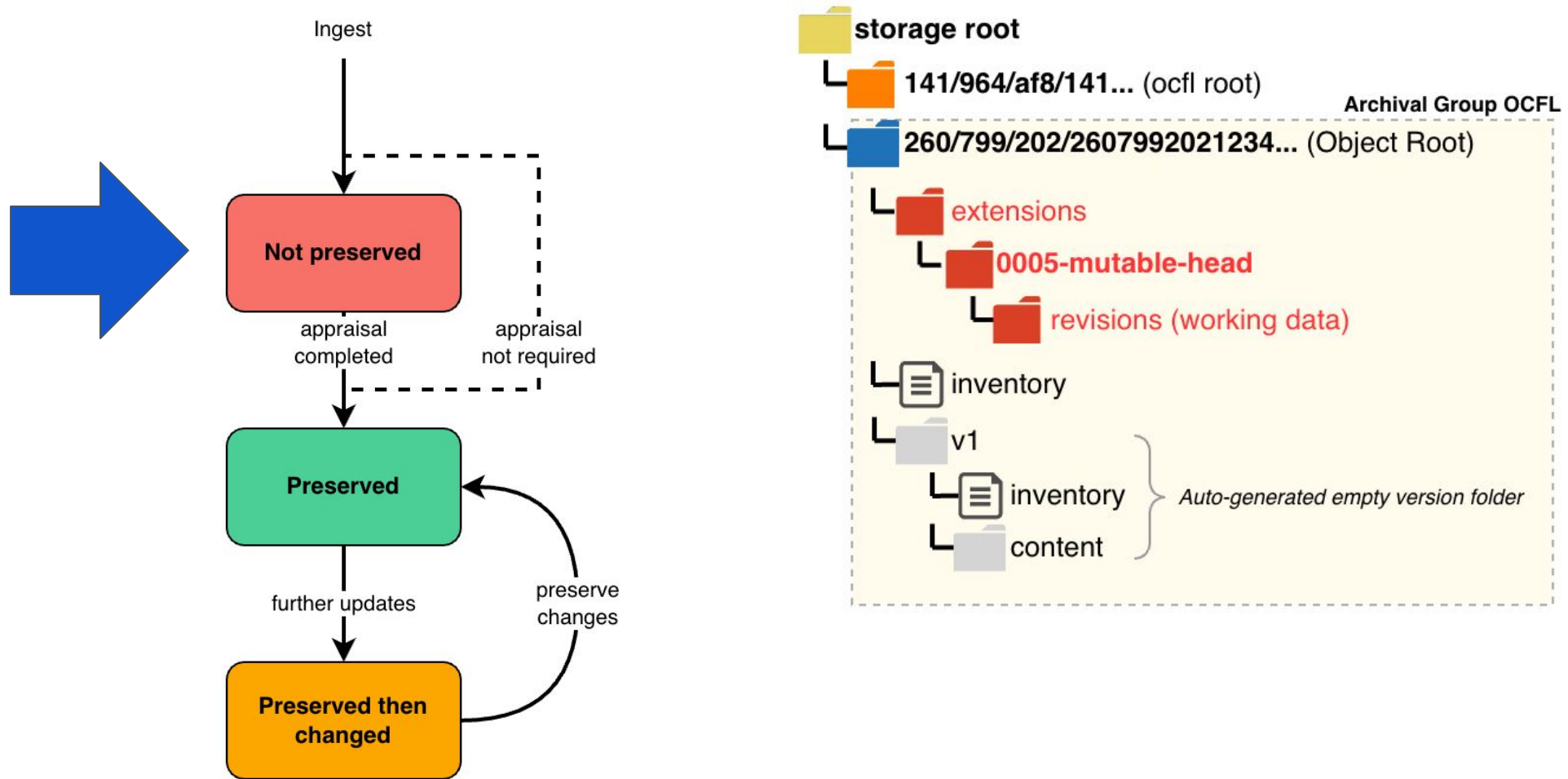
Find out more on our blog:



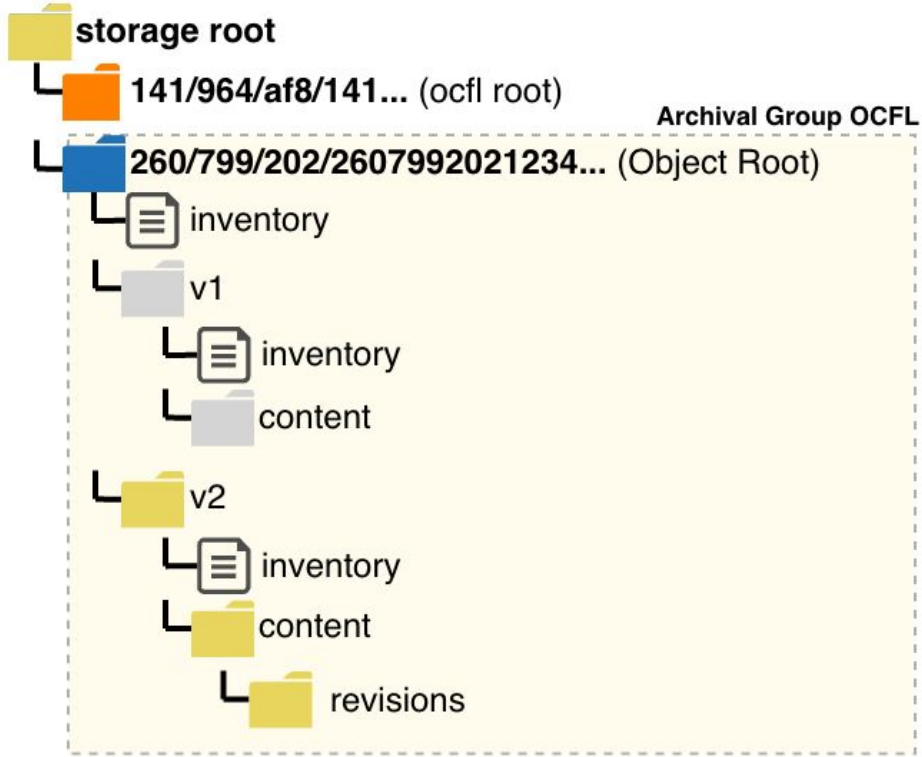
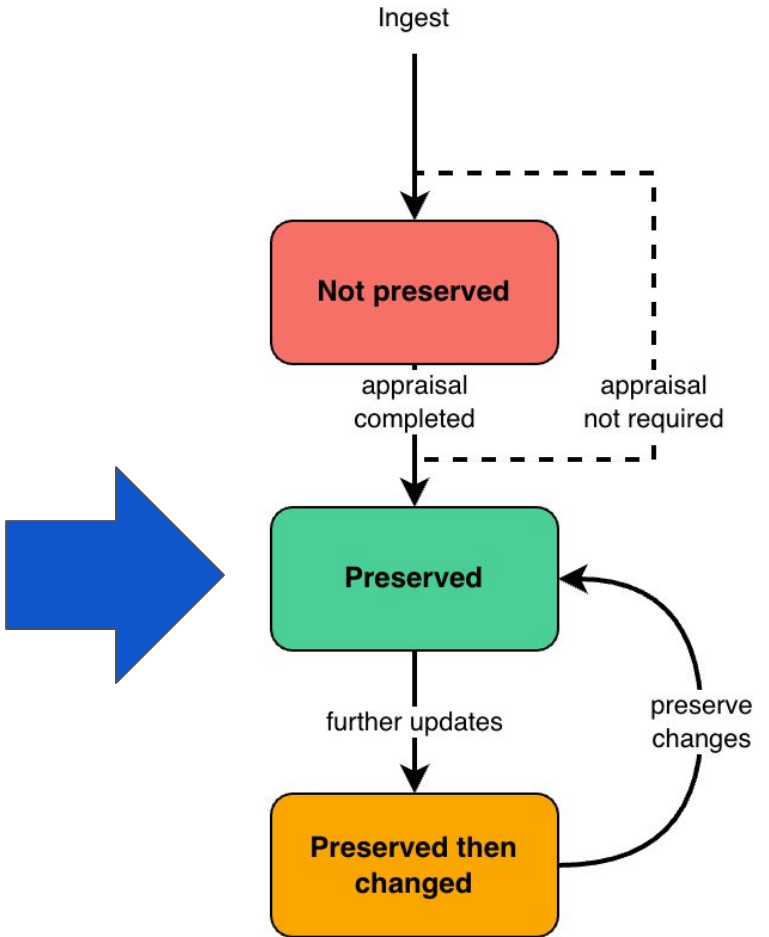
Ingest Workflows (Current & Planned)



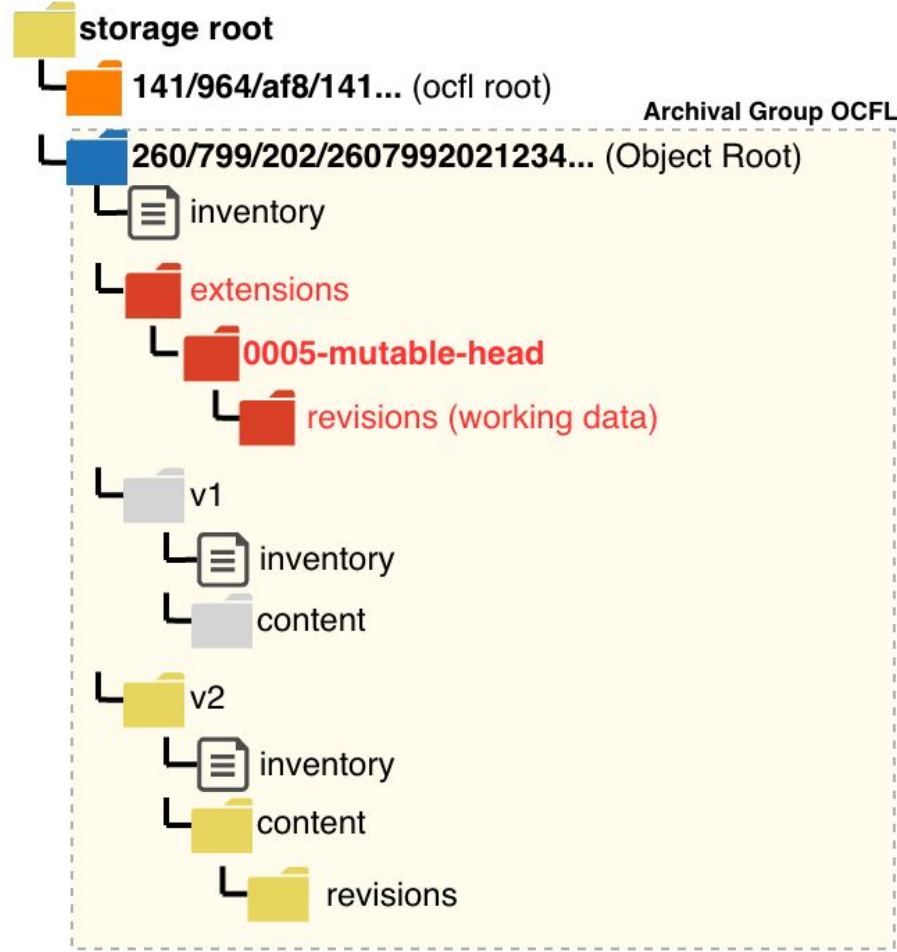
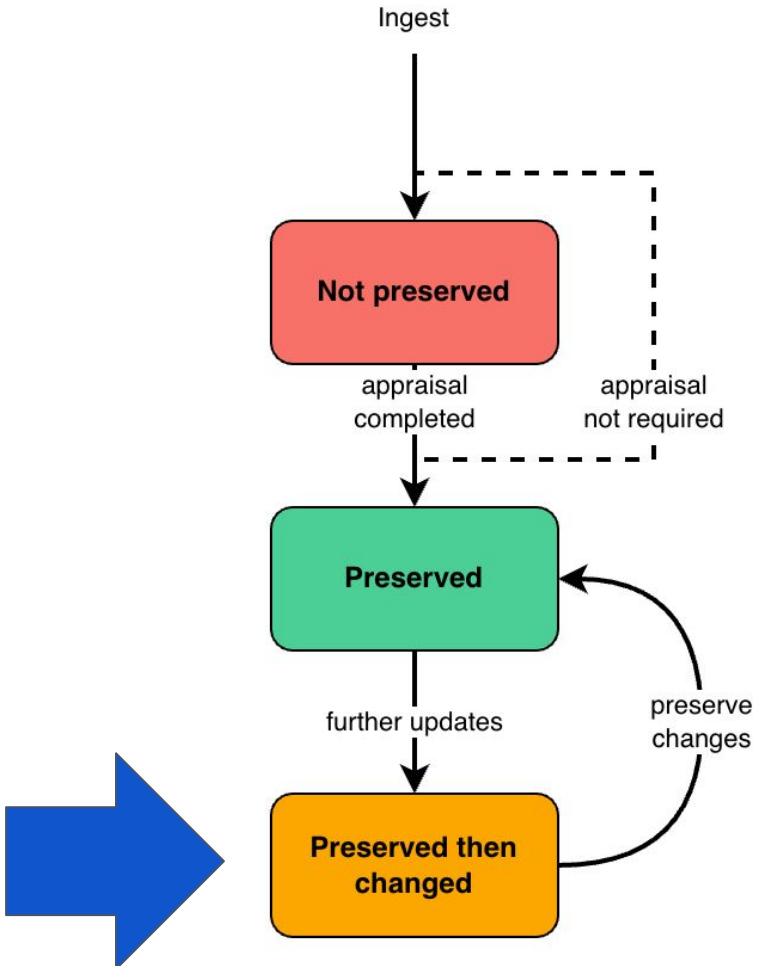
Storage Lifecycle & OCFL Structure



Storage Lifecycle & OCFL Structure



Storage Lifecycle & OCFL Structure

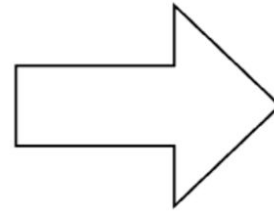
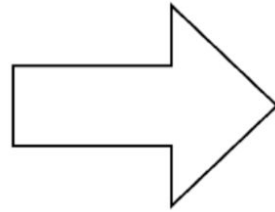


Copying to preservation storage

1st copy

2nd copy

3rd copy



**Amazon Web Services
(AWS)**
S3 Bucket

**Amazon Web Services
(AWS)**
S3 Bucket

Microsoft Azure
Blob Container

Repository and Preservation

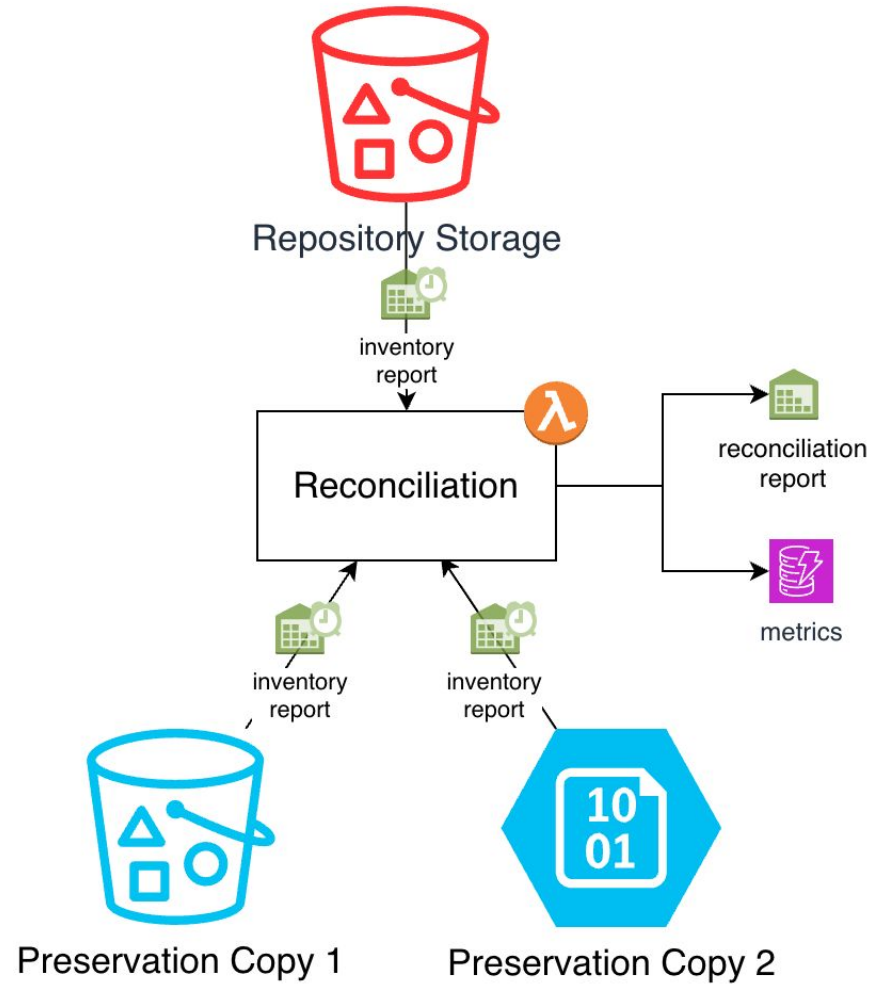
Preservation only

Preservation only

Chain of integrity

- File checksums calculated/captured at source:
 - during upload to Deposit Service
 - as part of transfer from carrier process
 - supplied in metadata
- All known checksums passed to Fedora for verification on submission
- Fedora stores these checksums in triples in metadata in OCFL
- sha256 passed as Additional Checksum to AWS for verification when copying to preservation
- md5 calculated at the same time (or use existing ETAG)
- md5 passed to Azure for verification when copying to preservation

Fixity & Reconciliation



Workbench UI: Features

- Monitor workflows (ingest & preservation)
- Search
- View/Download
- Appraisal & Accessioning
 - Edit/update metadata
 - Soft delete files
 - Flag system & empty files
 - Flag duplicates
 - Export metadata
 - Notes & tags
 - Preserve
 - Audit log of changes
- Reports
- User management

The screenshot displays the Digital Preservation Service Workbench interface. The header shows the application name 'Digital Preservation Service Workbench' in 'Development' mode, with the user 'J.A. GOSTICK' logged in. The left sidebar contains a navigation menu with options: Dashboard, Pre-Ingest, Workflows (with sub-items for Ingest (eTheses), Ingest, and Appraisals), Appraise (with sub-items for All Items, All Files, and My Items), Search, Reports, and Settings. The main content area shows a search result for 'Records of the CU Cruising Club 2022-3'. It includes action buttons for 'ASSIGN TO ME', 'VIEW ACCESSION IN ARCHIVSPACE', 'EDIT', and 'EXPORT METADATA'. Below these are creation and modification timestamps. A table lists metadata fields such as Triple Name, Fedora ID, Accession Number, Appraisal Required, ArchivesSpace Accession Id, Changed Since Last Version, Collecting Area, and Collecting Body. At the bottom, there are links for 'User Documentation', 'Send Feedback', and 'Report a Problem'.

Search (Items) / Records of the CU Cruising Club 2022-3

Records of the CU Cruising Club 2022-3

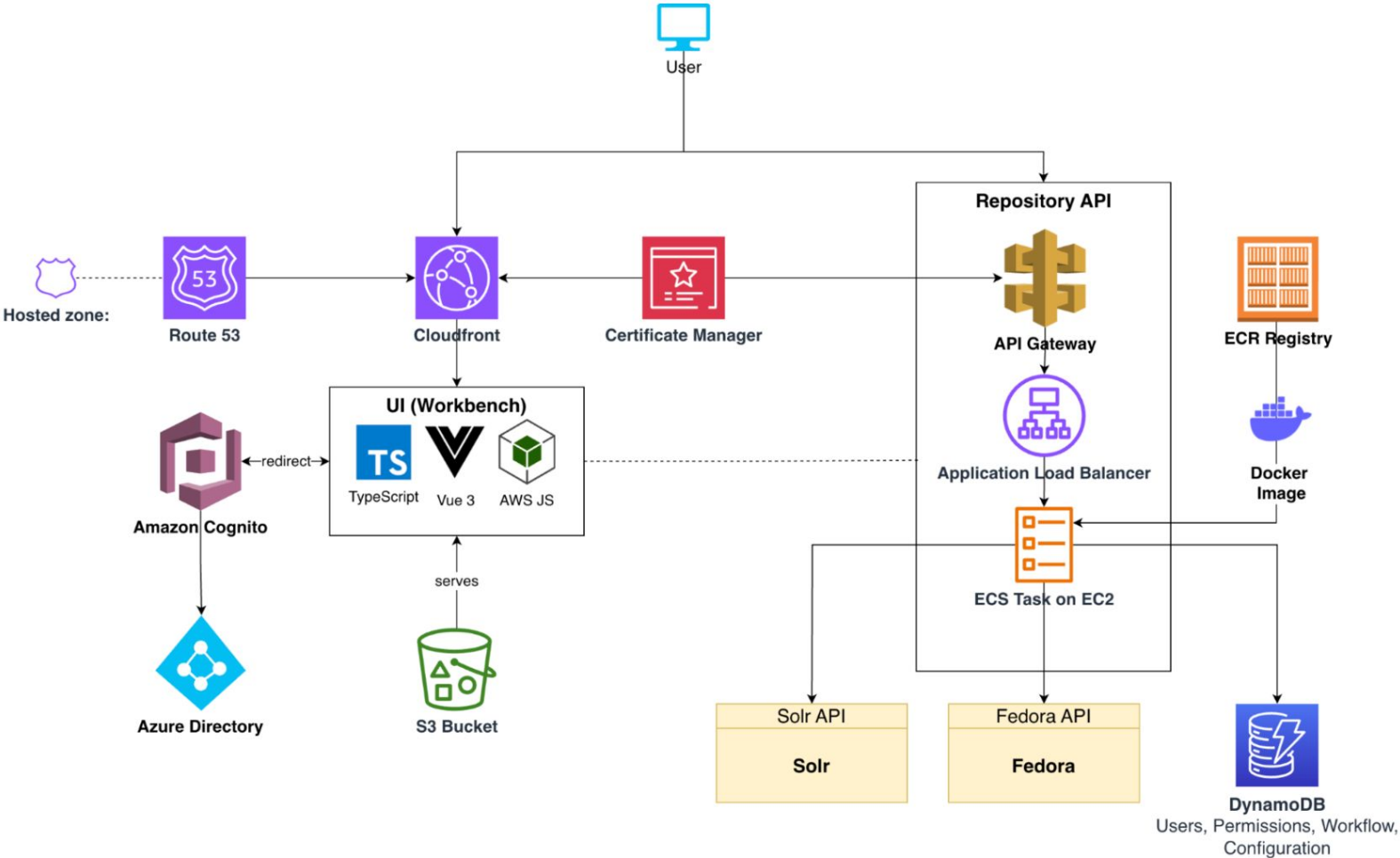
ASSIGN TO ME VIEW ACCESSION IN ARCHIVSPACE EDIT EXPORT METADATA

Created: 2025-02-14T17:04:02.242167Z by fedoraAdmin Last modified: 2025-03-17T16:14:39.827221Z by fedoraAdmin

Triple Name	Triple Value
Fedora ID	http://fcrepo.fedora:8080/fcrepo/rest/6d9becb0-41f2-432f-8
Accession Number	
Appraisal Required	true
ArchivesSpace Accession Id	456
Changed Since Last Version	true
Collecting Area	Department of Archives and Mss
Collecting Body	Cambridge University Library

For help, check out the [User Documentation](#) | [Send Feedback](#) or [Report a Problem](#)

Workbench & Repository API: AWS Architecture





We're hiring!

Senior Software Developer (Fixed Term)

<https://www.jobs.cam.ac.uk/job/50773/>

Thank you!

Any questions?

iag245@cam.ac.uk

tp425@cam.ac.uk

digitalpreservation@lib.cam.ac.uk

Visit our blog!

<https://digitalpreservation-blog.lib.cam.ac.uk/>

