Digital Preservation Infrastructure at Yale University Library

Euan Cochrane

Digital Preservation Manager

euan.cochrane@yale.edu

March 2022

https://twitter.com/euanc

1.3 Petabytes+ of
Data Ingested

120+ Million files

~1.6TB of
Metadata

Preservica
Active digital preservation

Our Users

Yale University Library Tenancy

- Manuscripts and Archives (MSSA)
- Beinecke Rare Book and Manuscripts Library (BRBL)
- Divinity Library
- Irving S. Gilmore Music Library
- Lewis Walpole Library
- Library IT
- Marx Science and Social Science Library
- Harvey Cushing/John Hay Whitney Medical Library
- Fortunoff Holocaust Testimonies Archive
- Robert B. Haas Family Arts Library
- General Collections
- Digital Humanities

Museums Tenancy

- Yale Center for British Art
- Yale Art Gallery
- Peabody Museum

Law Library Tenancy

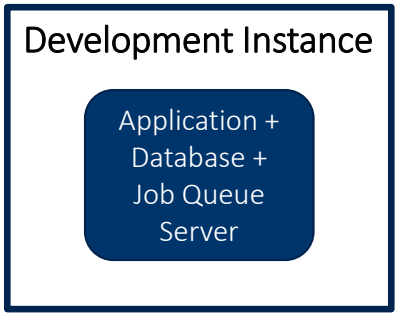
- Lillian Goldman Law Library

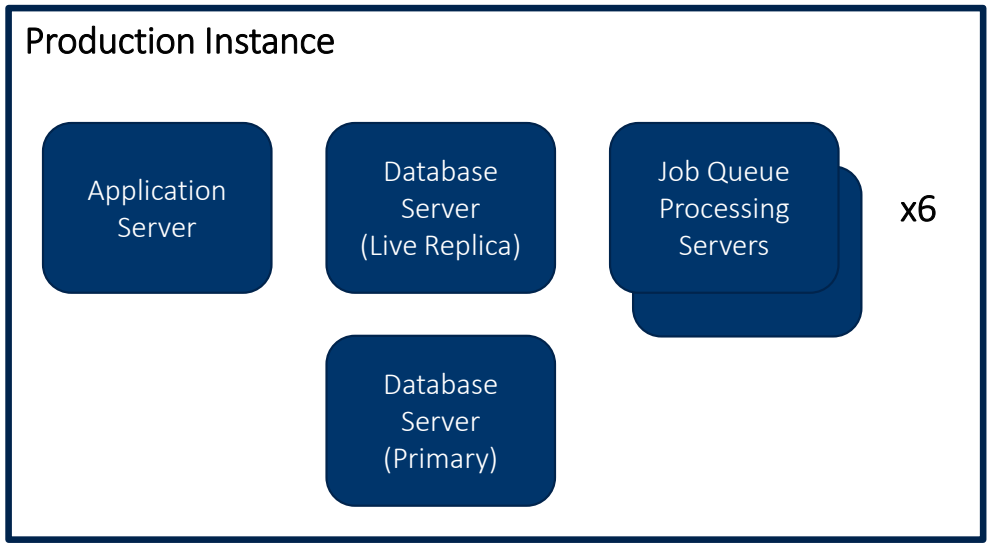
How we work with our users

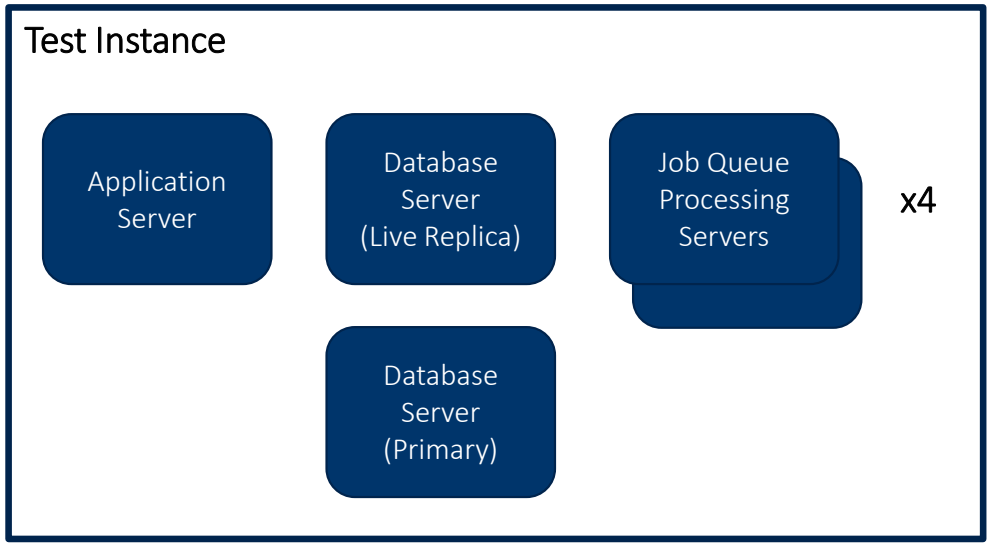
1. Initial meetings to discuss how they want to work with Preservica and our team
2. Users decide risks and storage options based on published policy
3. Users decide security model
4. We train users
5. For big ingests we may help automate them
6. Staff in units ingest content themselves
7. We undertake migrations on their behalf
8. We develop reports for units
9. We check in regularly for planning for large ingests (there is a lead time for some of the storage services we use) and to make sure they are getting everything they need
10. Ticketing via a ticketing service
11. System upgrades at least once per year

Infrastructure

VirtualMachine Instances







Storage Options

We use available services

1. Local Tape Storage*
   1. Replicated
2. Local Disk Storage
   1. Replicated
3. Cloud - Amazon Glacier
4. Cloud - Wasabi

Explorations:

1. DNA Storage
   Pilot underway with Twist BioScience
2. PIQL Film Storage
   Pilot underway

*Was provided by Yale ITS, currently Library IT are in the process of implementing a replacement

Digital Preservation Storage Policy

Principle 1: Maximize the number of copies of stored content

Principle 2: Maximize the diversity of storage systems, administrators and technologies in use

Principle 3: Select lower cost options (all else being equal)

#	Content type	Value	Relative volume	Relative retrieval frequency	Storage Recommendation	# of copies	Relative risk Level	Relative cost per GB
1	Born Digital Masters	High	Low	Medium	Disk + Tape + Cloud	6+	Low	High
2	Digitized Masters (A/V)	High	High	Low	Tape + Cloud	4+	Medium-low	Medium
3	Digitized Masters (Textual)	High	medium	Medium	Disk + Tape + Cloud	6+	Low	High
4	Digitized Surrogate Masters (Textual and A/V)	Low	High	Low	Tape	2	Medium	Very low
5	Access Derivatives	Very Low	High	High	Disk	2	Medium	Low

https://bit.ly/YaleStoragePolicy

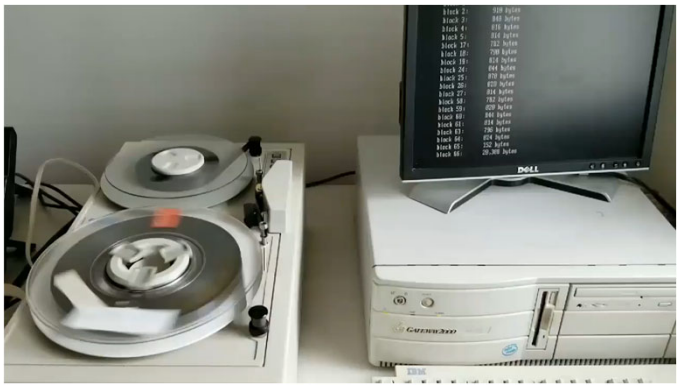
Storage in Preservica

1. Storage locations ("adapters) configured per type, per funder
2. Rules configured for routing content based on information in metadata, e.g. source, type
3. Rules applied on ingest (or when a storage update workflow is run)
4. Data automatically replicated by Preservica/storage services
5. Preservica configured to check regularly storage options where possible.

Data Recovery Services

We work with:

- Special collections material out of scope of the digital accessioning service
- Faculty on an ad-hoc basis
- General collections
  - E.g. Acquisitions include publications on microSD cards



The Digital Preservation System Team

Preservation

David Cirella – Digital Preservation Librarian

Euan Cochrane – Digital Preservation Manager

Grete Graf – Digital Preservation Librarian

Library IT

Software Developer - we "fund" capacity

System Administrators (Devops and Infrastructure)

Bob Rice – Technical Team Lead

Keith Boyd-Carter – Operating Systems Programmer

Ricardo Aliwalis – Linux Administrator

Thank you

euan.cochrane@yale.edu

robert.rice@yale.edu

https://bit.ly/YaleStoragePolicy