PRESENT AND POSSIBLE FUTURES OF DIGITAL PRESERVATION AT CLEMSON LIBRARIES

KELLY RIDDLE

3 FEBRUARY 2025

DIGITAL PRESERVATION COALITION
WORKFLOW WEBINAR

Outline

- About Clemson, Clemson Libraries, and Digitization Services
- About Our Content
- Current Digital Preservation Workflow
- Digital Preservation Assessments
- Possible Future Digital Preservation Workflow
- Next Steps

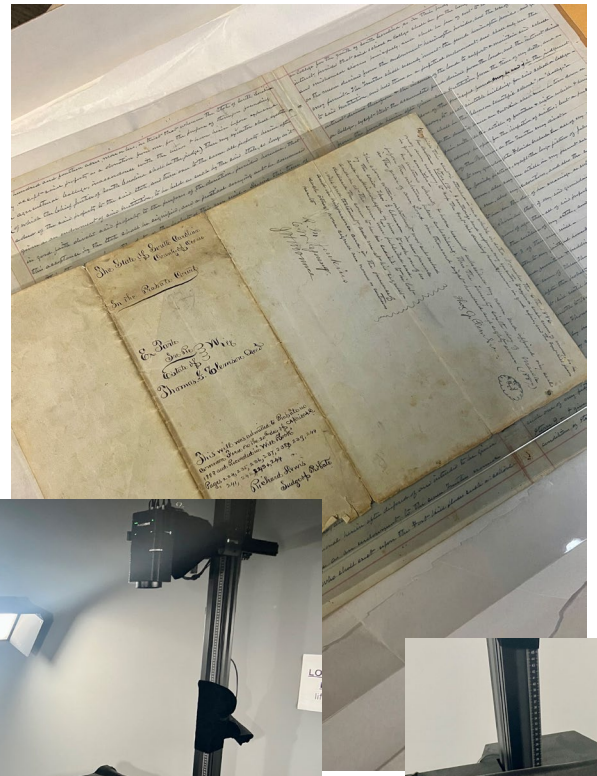
About Clemson

- Public land-grant university in South Carolina, USA
- 28,747 enrolled students (Fall 2023)
- 1,605 Faculty
- Carnegie R1
- Main campus in Clemson
- Satellite campus in Greenville
- Cooperative Extension Service
- STEM-based curriculum

About Clemson Libraries

- Robert Muldrow Cooper Library
- Education Media Center
- Emery A. Gunnin Architecture Library
- Special Collections and Archives
- Library Depot –Digital Imaging Lab
- Historic Properties
- University Press – ClemsonOpen
  Institutional Repository

About Digitization Services



2016

- Current rackstation servers brought online for digital projects storage
- Current digital collections sites launched

2024

Current Director onboarded

2007

Digitization Services founded in Office of Library Technology

2022

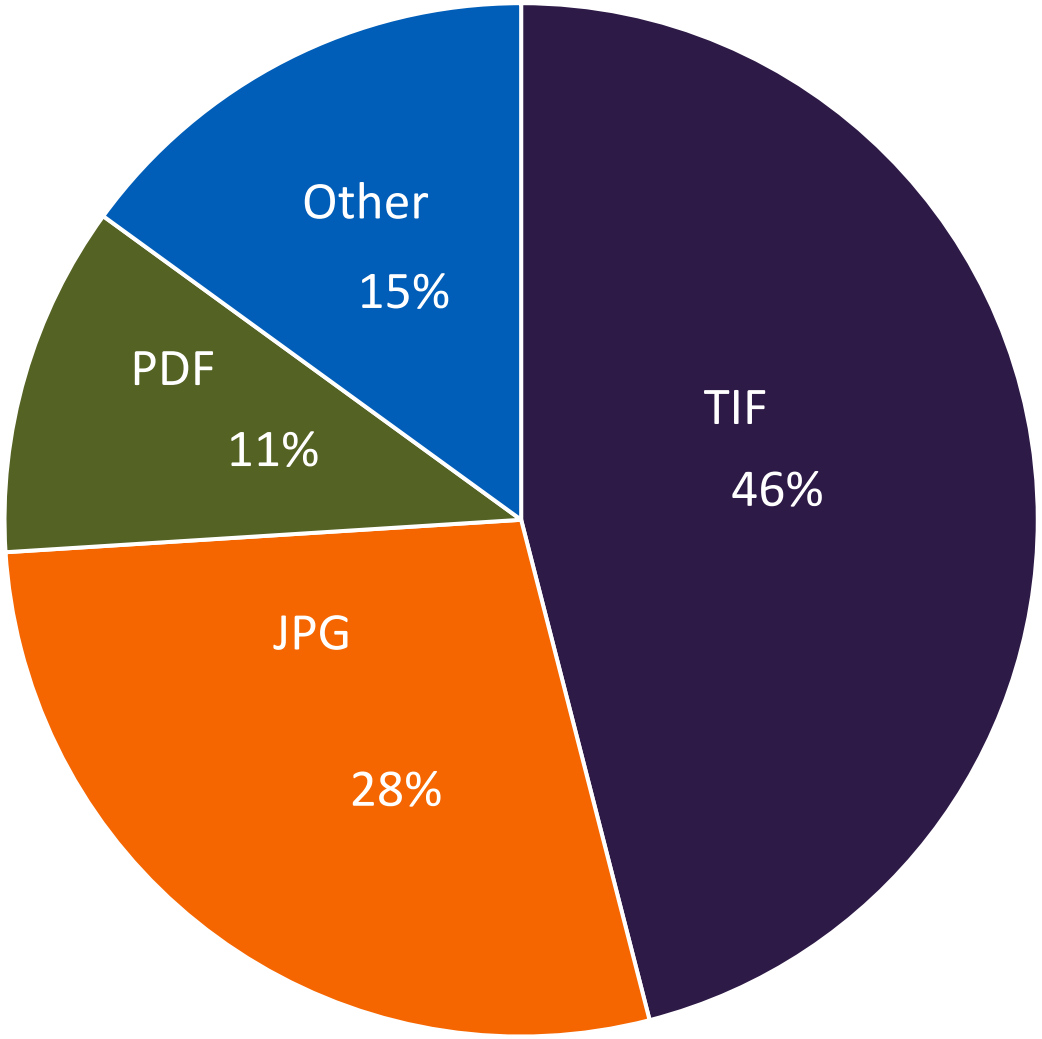
Digitization Services moved from Office of Library Technology to Division of Collections and Discovery

About Our Content

- Process
  - Imaging workflows in Digital Imaging Lab
- Significant Digital Collections
  - Clemson Special Collections and Archives
  - Open Parks Network
  - Theses and Dissertations
  - South Carolina Digital Library
- Age
  - Digitization started in 2007
  - Current storage infrastructure brought online in 2016

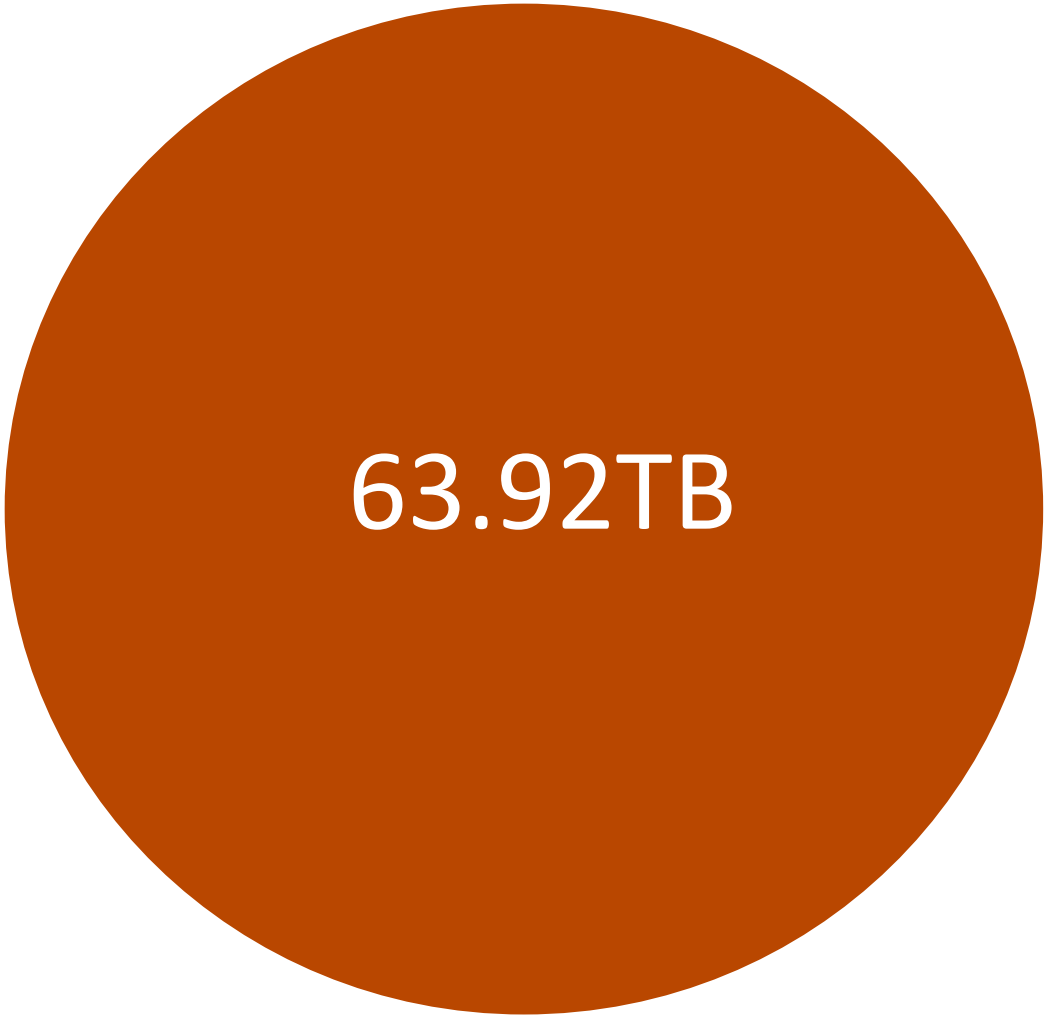
About Our Content: Extent and Distribution

- Collections: 222
- Unique File Types: 2,523
- Total Files: 3,003,058
  - TIF 1,388,806
  - JPG 859,155
  - PDF 350,104
  - TXT 11,3989
  - JP2 10,6076
  - WAV 3,065
  - MP4 1,644

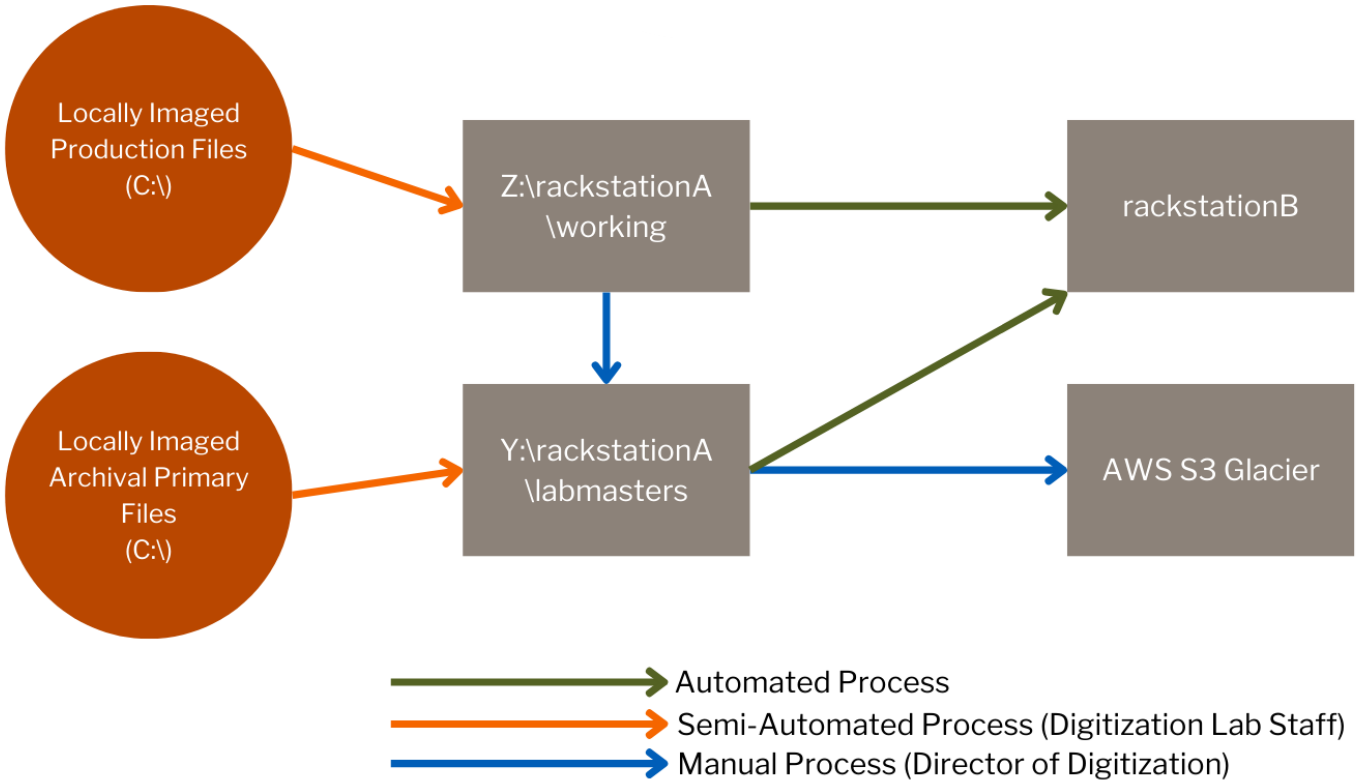


About Our Content: Extent and Distribution

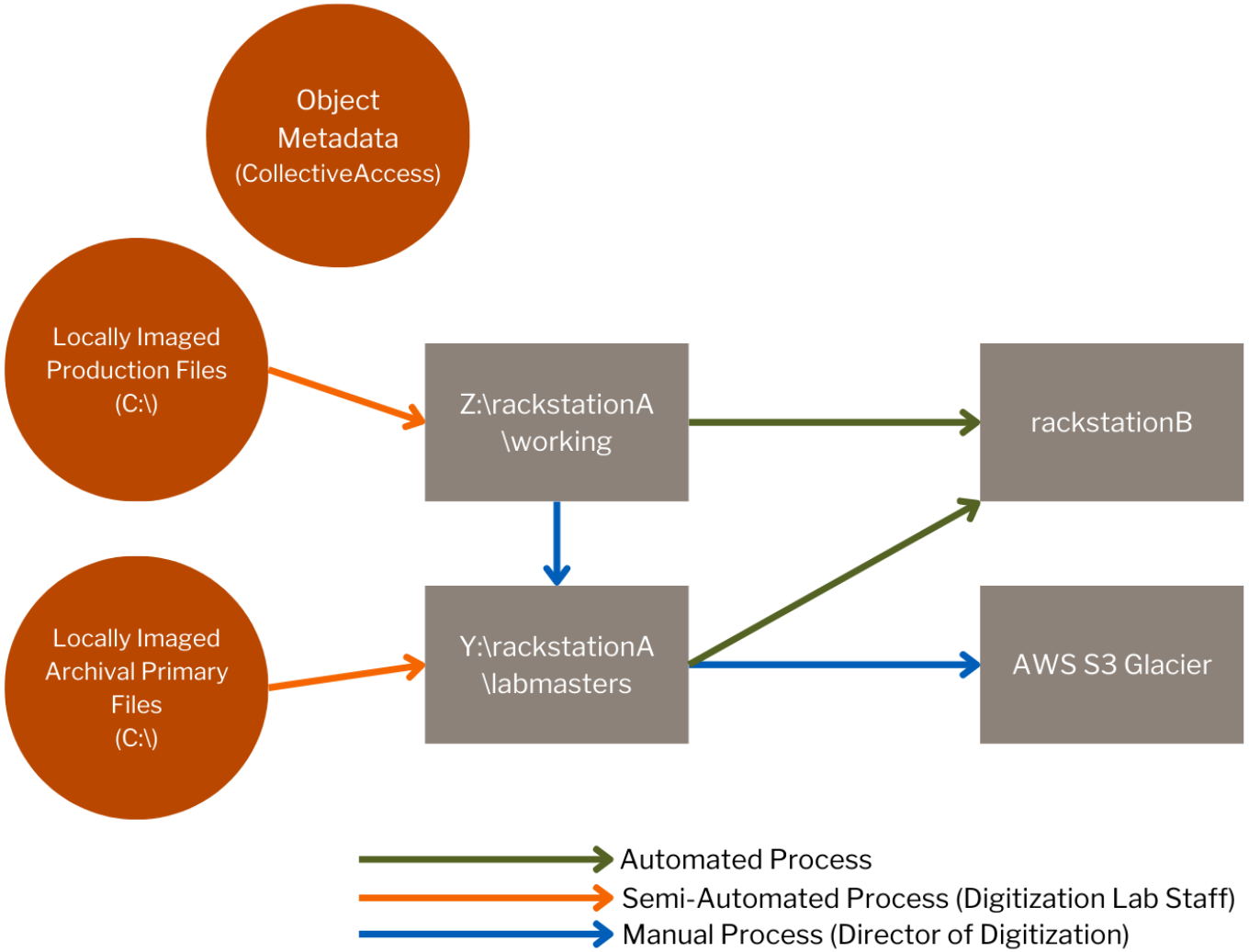
- Collections: 222
- Unique File Types: 2,523
- Total Files: 3,003,058
  - TIF 1,388,806
  - JPG 859,155
  - PDF 350,104
  - TXT 11,3989
  - JP2 10,6076
  - WAV 3,065
  - MP4 1,644



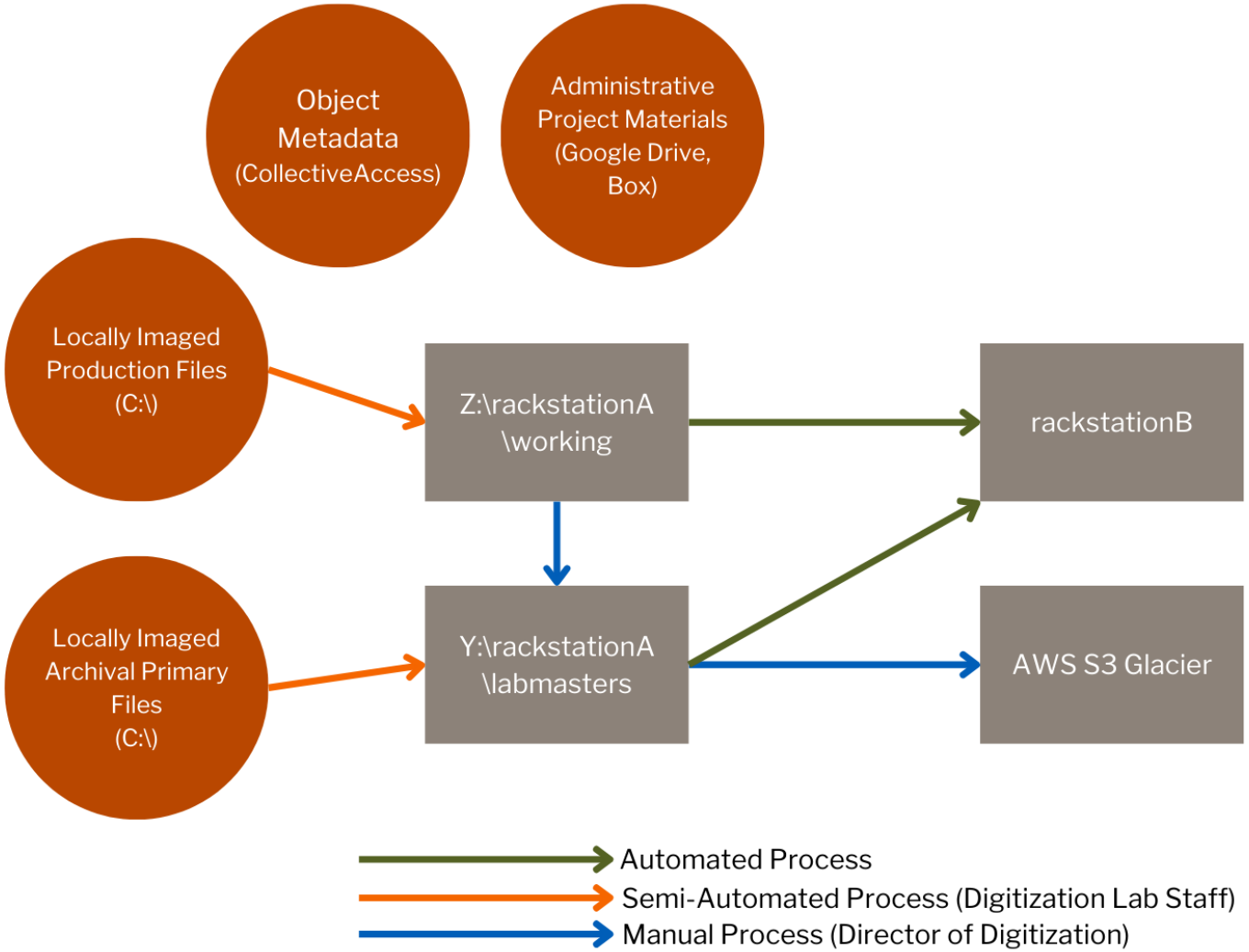
Current Digital Preservation Workflow



Current Digital Preservation Workflow



Current Digital Preservation Workflow



Assessments

Digital Preservation Coalition Rapid Assessment Model

National Digital Stewardship Alliance Levels of Digital Preservation

NDSA

Assessment – DPC Rapid Assessment Model

ORGANIZATIONAL CAPABILITIES	
	Current Level
A. Organizational viability: Governance, organizational structure, staffing and resourcing of digital preservation activities.	
B. Policy and strategy: Policies, strategies, and procedures which govern the operation and management of the digital archive.	
C. Legal and ethical: Management of legal, social and cultural rights and responsibilities, compliance with relevant regulation and adherence to codes of ethics related to acquiring, preserving and providing access to digital content.	
D. IT capability: Information Technology capabilities for supporting digital preservation activities.	
E. Continuous Improvement: Processes for the assessment of current digital preservation capabilities, the definition of goals and the monitoring of progress	
F. Community: Engagement with and contribution to the wider digital preservation community.	

SERVICE CAPABILITIES	
	Current Level
G. Acquisition, Transfer and Ingest: Processes to acquire or transfer content and ingest it into a digital archive.	
H. Bitstream Preservation: Processes to ensure the storage and integrity of digital content to be preserved.	
I. Content Preservation: Processes to preserve the meaning, usability and functionality of the digital content over time.	
J. Metadata Management: Processes to create and maintain sufficient metadata to support preservation, discovery and use of preserved digital content.	
K. Discovery and Access: Processes to enable discovery of digital content and provide access for users.	

Assessment – DPC Rapid Assessment Model

ORGANIZATIONAL CAPABILITIES	
	Current Level
A. Organizational viability: Governance, organizational structure, staffing and resourcing of digital preservation activities.	1 - Awareness
B. Policy and strategy: Policies, strategies, and procedures which govern the operation and management of the digital archive.	0 - Minimal Awareness
C. Legal and ethical: Management of legal, social and cultural rights and responsibilities, compliance with relevant regulation and adherence to codes of ethics related to acquiring, preserving and providing access to digital content.	1 - Awareness
D. IT capability: Information Technology capabilities for supporting digital preservation activities.	3 - Managed
E. Continuous Improvement: Processes for the assessment of current digital preservation capabilities, the definition of goals and the monitoring of progress	2 - Basic
F. Community: Engagement with and contribution to the wider digital preservation community.	2 - Basic

SERVICE CAPABILITIES	
	Current Level
G. Acquisition, Transfer and Ingest: Processes to acquire or transfer content and ingest it into a digital archive.	1 - Awareness
H. Bitstream Preservation: Processes to ensure the storage and integrity of digital content to be preserved.	1 - Awareness
I. Content Preservation: Processes to preserve the meaning, usability and functionality of the digital content over time.	1 - Awareness
J. Metadata Management: Processes to create and maintain sufficient metadata to support preservation, discovery and use of preserved digital content.	1 - Awareness
K. Discovery and Access: Processes to enable discovery of digital content and provide access for users.	2 - Basic

Levels
0 – Minimal awareness
1 – Awareness
2 – Basic
3 – Managed
4 – Optimized

Assessment – NDSA Levels

NDSA

Levels of Digital Preservation

Functional Area	Level			
	Level 1 (Know your content)	**Level 2 (Protect your content)**	**Level 3 (Monitor your content)**	**Level 4 (Sustain your content)**
Storage	Have two complete copies in separate locations Document all storage media where content is stored Put content into stable storage	Have three complete copies with at least one copy in a separate geographic location Document storage and storage media indicating the resources and dependencies they require to function	Have at least one copy in a geographic location with a different disaster threat than the other copies Have at least one copy on a different storage media type Track the obsolescence of storage and media	Have at least three copies in geographic locations, each with a different disaster threat Maximize storage diversification to avoid single points of failure Have a plan and execute actions to address obsolescence of storage hardware, software, and media
Integrity	Verify integrity information if it has been provided with the content Generate integrity information if not provided with the content Virus check all content; isolate content for quarantine as needed	Verify integrity information when moving or copying content Use write-blockers when working with original media Back up integrity information and store copy in a separate location from the content	Verify integrity information of content at fixed intervals Document integrity information verification processes and outcomes Perform audit of integrity information on demand	Verify integrity information in response to specific events or activities Replace or repair corrupted content as necessary
Control	Determine the human and software agents that should be authorized to read, write, move, and delete content	Document the human and software agents authorized to read, write, move, and delete content and apply these	Maintain logs and identify the human and software agents that performed actions on content	Perform periodic review of actions/access logs
Metadata	Create inventory of content, also documenting current storage locations Backup inventory and store at least one copy separately from content	Store enough metadata to know what the content is (this might include some combination of administrative, technical, descriptive, preservation, and structural)	Determine what metadata standards to apply Find and fill gaps in your metadata to meet those standards	Record preservation actions associated with content and when those actions occur Implement metadata standards chosen
Content	Document file formats and other essential content characteristics including how and when these were identified	Verify file formats and other essential content characteristics Build relationships with content creators to encourage sustainable file choices	Monitor for obsolescence, and changes in technologies on which content is dependent	Perform migrations, normalizations, emulation, and similar activities that ensure content can be accessed

Levels of Digital Preservation v2.0

Assessment – NDSA Levels

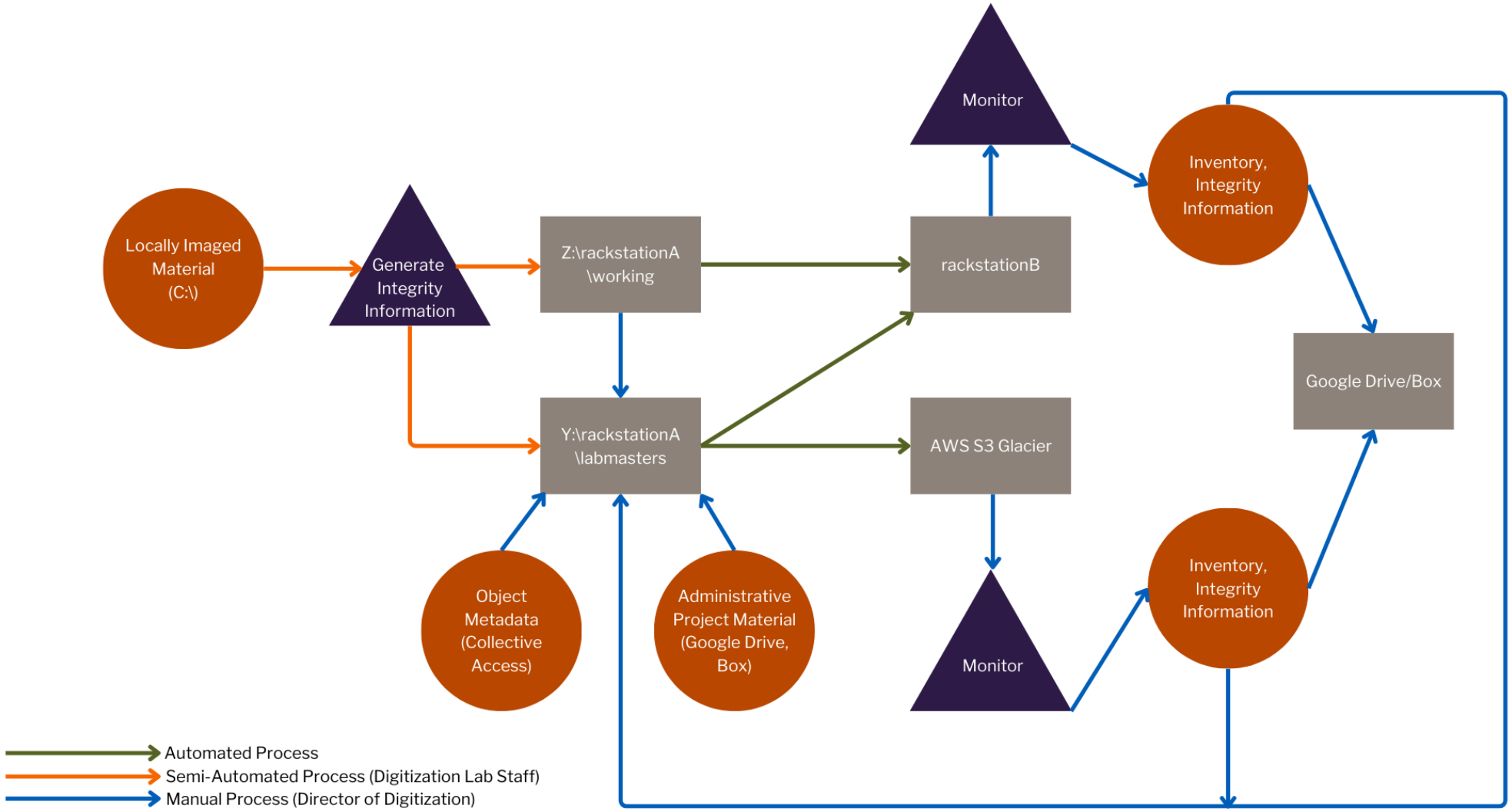
NDSA

Levels of Digital Preservation

Functional Area	Level			
	Level 1 (Know your content)	**Level 2 (Protect your content)**	**Level 3 (Monitor your content)**	**Level 4 (Sustain your content)**
Storage	Have two complete copies in separate locations Document all storage media where content is stored Put content into stable storage	Have three complete copies with at least one copy in a separate geographic location Document storage and storage media indicating the resources and dependencies they require to function	Have at least one copy in a geographic location with a different disaster threat than the other copies Have at least one copy on a different storage media type Track the obsolescence of storage and media	Have at least three copies in geographic locations, each with a different disaster threat Maximize storage diversification to avoid single points of failure Have a plan and execute actions to address obsolescence of storage hardware, software, and media
Integrity	Verify integrity information if it has been provided with the content Generate integrity information if not provided with the content Virus check all content; isolate content for quarantine as needed	Verify integrity information when moving or copying content Use write-blockers when working with original media Back up integrity information and store copy in a separate location from the content	Verify integrity information of content at fixed intervals Document integrity information verification processes and outcomes Perform audit of integrity information on demand	Verify integrity information in response to specific events or activities Replace or repair corrupted content as necessary
Control	Determine the human and software agents that should be authorized to read, write, move, and delete content	Document the human and software agents authorized to read, write, move, and delete content and apply these	Maintain logs and identify the human and software agents that performed actions on content	Perform periodic review of actions/access logs
Metadata	Create inventory of content, also documenting current storage locations Backup inventory and store at least one copy separately from content	Store enough metadata to know what the content is (this might include some combination of administrative, technical, descriptive, preservation, and structural)	Determine what metadata standards to apply Find and fill gaps in your metadata to meet those standards	Record preservation actions associated with content and when those actions occur Implement metadata standards chosen
Content	Document file formats and other essential content characteristics including how and when these were identified	Verify file formats and other essential content characteristics Build relationships with content creators to encourage sustainable file choices	Monitor for obsolescence, and changes in technologies on which content is dependent	Perform migrations, normalizations, emulation, and similar activities that ensure content can be accessed

Levels of Digital Preservation v2.0

Possible Future Digital Preservation Workflow



Next Steps

- Report on digital preservation assessments
- Convene Digital Projects Cross-Functional Team
- Relocate rackstationB from Cooper to Library Depot
- Implement new processes
- Plan for growth

References

Clemson University (2023) “Clemson Undergraduate Admission Statistics” https://www.clemson.edu/admissions/undergraduate-admissions/admissions-statistics.html.

Digital Preservation Coalition (2024) “Digital Preservation Coalition Rapid Assessment Model (DPC Ram) (version 3)” http://doi.org/10.7207/dpcram24-03.

Dutkiewicz, Scott M. and Serrao, Jessica L. (2019) "A Road Taken: A Cataloging Team Becomes a Metadata Team,” Publications. 158. https://tigerprints.clemson.edu/lib_pubs/158.

Levels of Preservation Revisions Working Group (2019) “Levels of Digital Preservation Matrix V2.0,” https://osf.io/2mkwx/.

R. M. Cooper Library (1966) https://digitalcollections.clemson.edu/single-item-view/?oid=CUIR:5700235C48F5913E167BC9FFB82A9A82&b=https%3A%2F%2Fdigitalcollections.clemson.edu%2Fsearch-results%2F%3Fk%3Dcooper%2520library%26opn_refine_control%3Djump%26xi%3D1%26xm%3D20%26sf%3D.

Tridgell, Andrew and Mackerras, Paul (2025) Rsynch Project (v3.4.1), https://github.com/RsyncProject/rsync.

QUESTIONS

KELLY RIDDLE
RIDDLE2@CLEMSON.EDU