ITHAKA

JSTOR | PORTICO | ITHAKA S+R

Protecting Future Access Now
Models for Preserving Locally Created Content

By Amy Kirchhoff
Archive Service Product Manager, Portico, ITHAKA

Amigos Online Conference
Digital Preservation: What's Now, What's Next
August 12, 2011

Amy Kirchhoff
Archive Service Product Manager
Portico, JSTOR

amy.kirchhoff@ithaka.org
609-986-2218

ITHAKA is a not-for-profit organization that helps the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways.

We pursue this mission by providing innovative services that aid in the adoption of these technologies and that create lasting impact.

Ithaka S+R is a research and consulting service that focuses on the transformation of scholarship and teaching in an online environment, with the goal of identifying the critical issues facing our community and acting as a catalyst for change.

JSTOR is a research platform that enables discovery, access, and preservation of scholarly content.

Portico is a digital preservation service for e-journals, e-books, and other scholarly e-content.

PORTICO

Over 2,000 societies, and associations have committed content to Portico through 132 publishers agreements.

- E-journal titles 12,204
- E-book titles 107,117
- D-collections 45

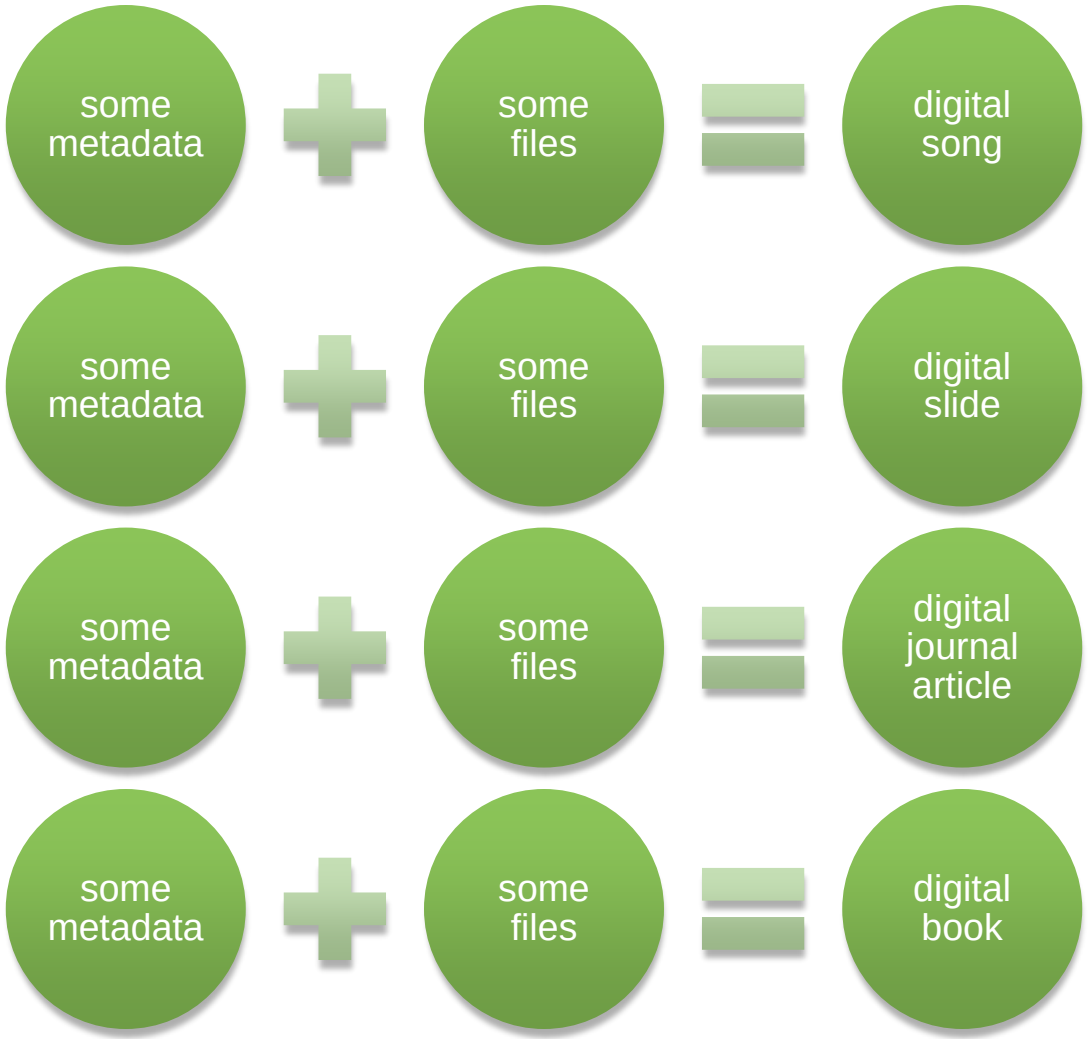
ANYONE CAN BE A PUBLISHER…

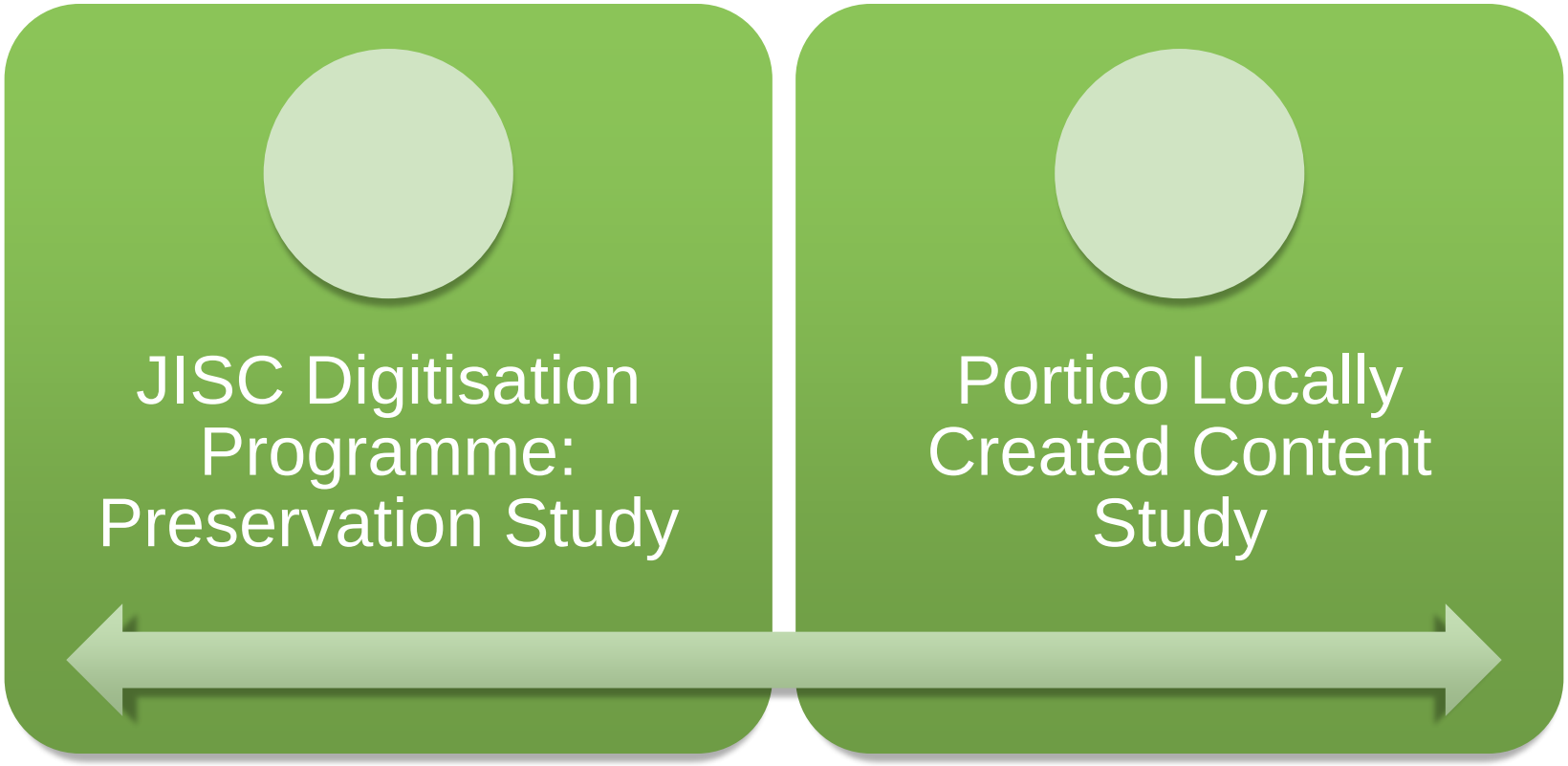
The NEH and IMLS awarded a grant to Portico, in partnership with Cornell University Library, through the "Advancing Knowledge: The IMLS/NEH Digital Partnership grant program" to develop a practical model for how preservation can be accomplished for digital content at cultural heritage organizations.

White Paper:

http://www.portico.org/presentations-publications/

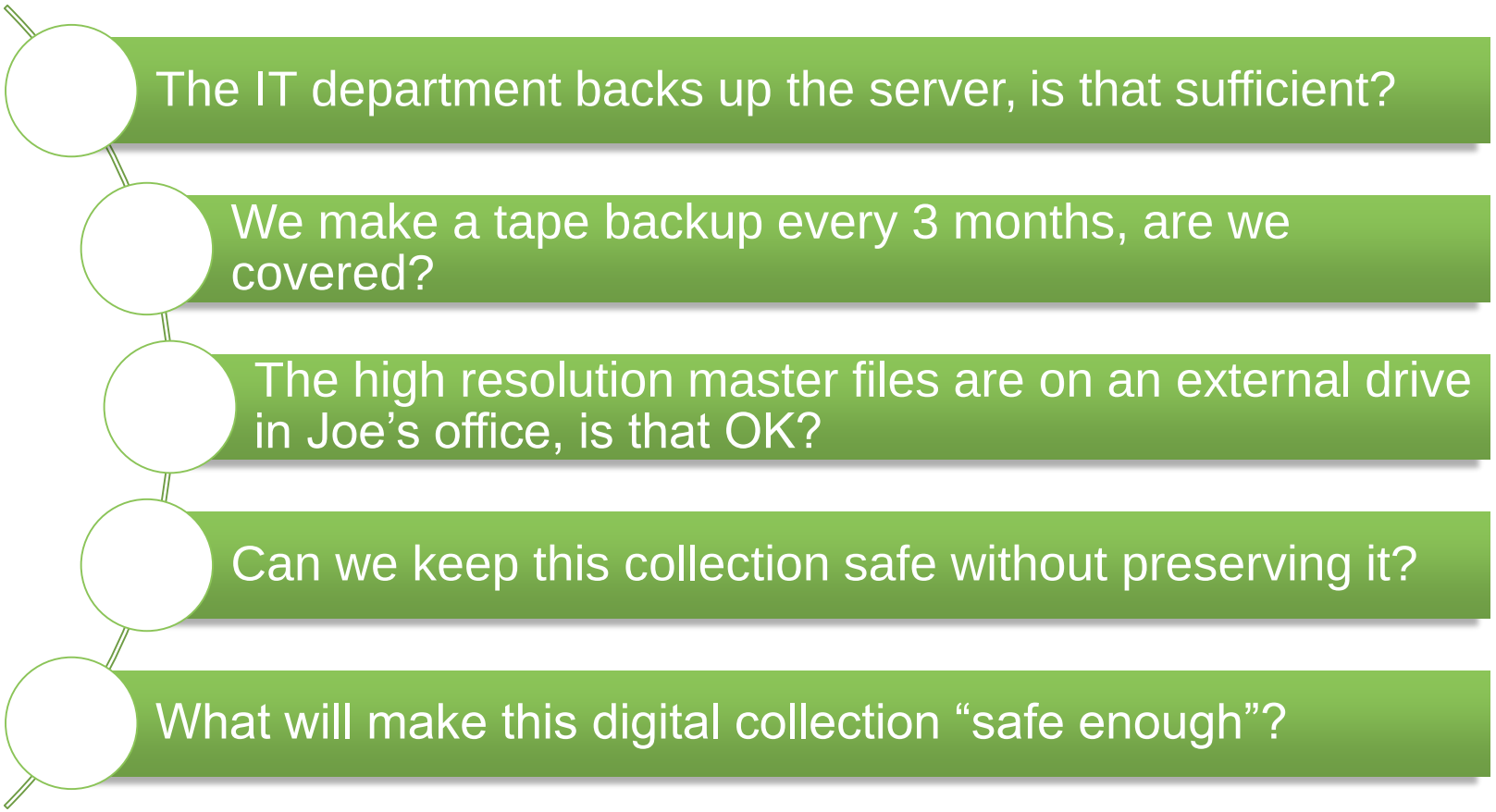
A Preservation Model





Many thanks to JISC and NEH for funding.

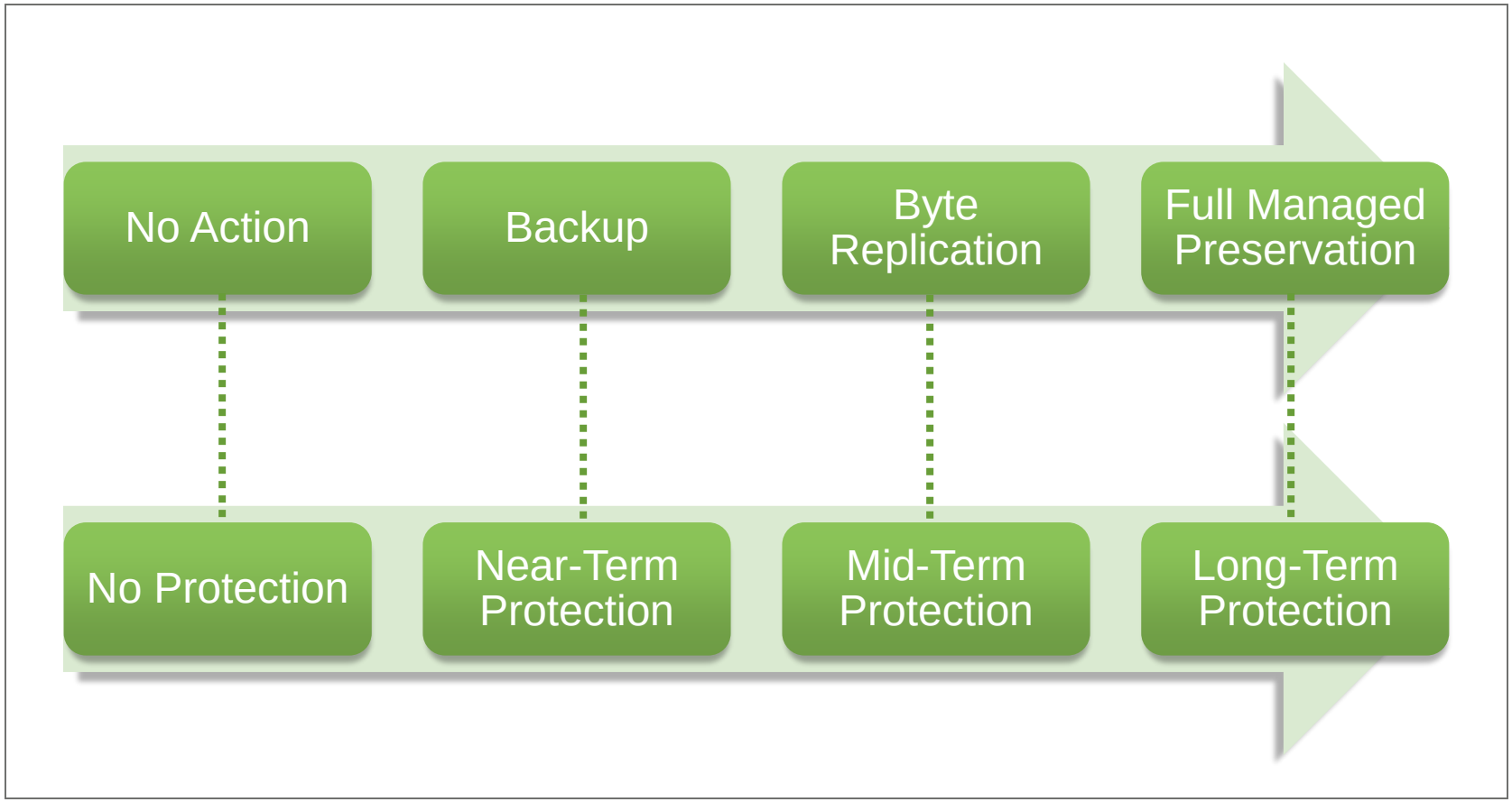
Reasonable Preservation Questions



Reasonable Preservation Questions

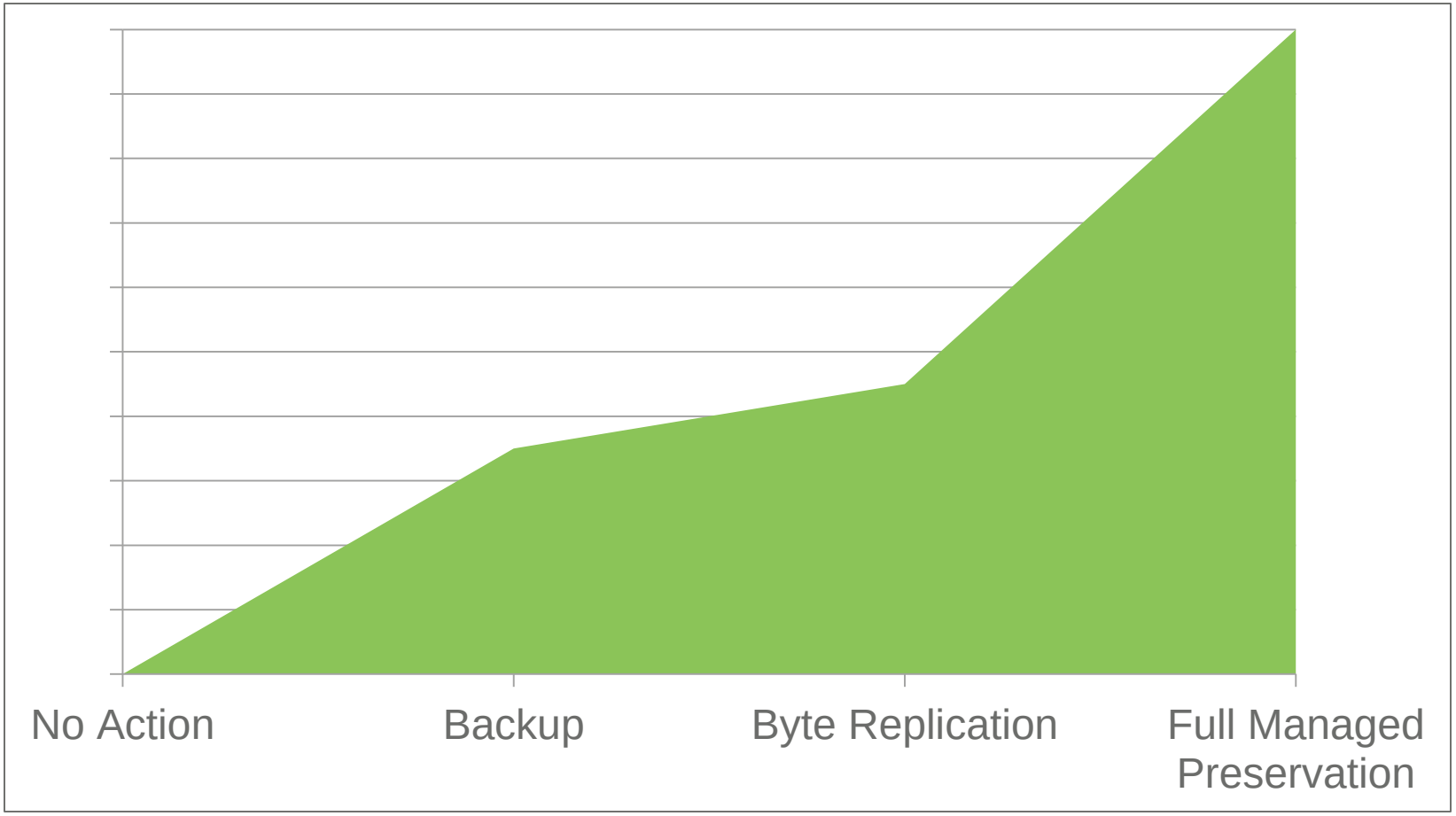
IT DEPENDS…

Reasonable Preservation Questions



CONTENT PROTECTION CONTINUUM

Reasonable Preservation Questions



CONTENT PROTECTION CONTINUUM II

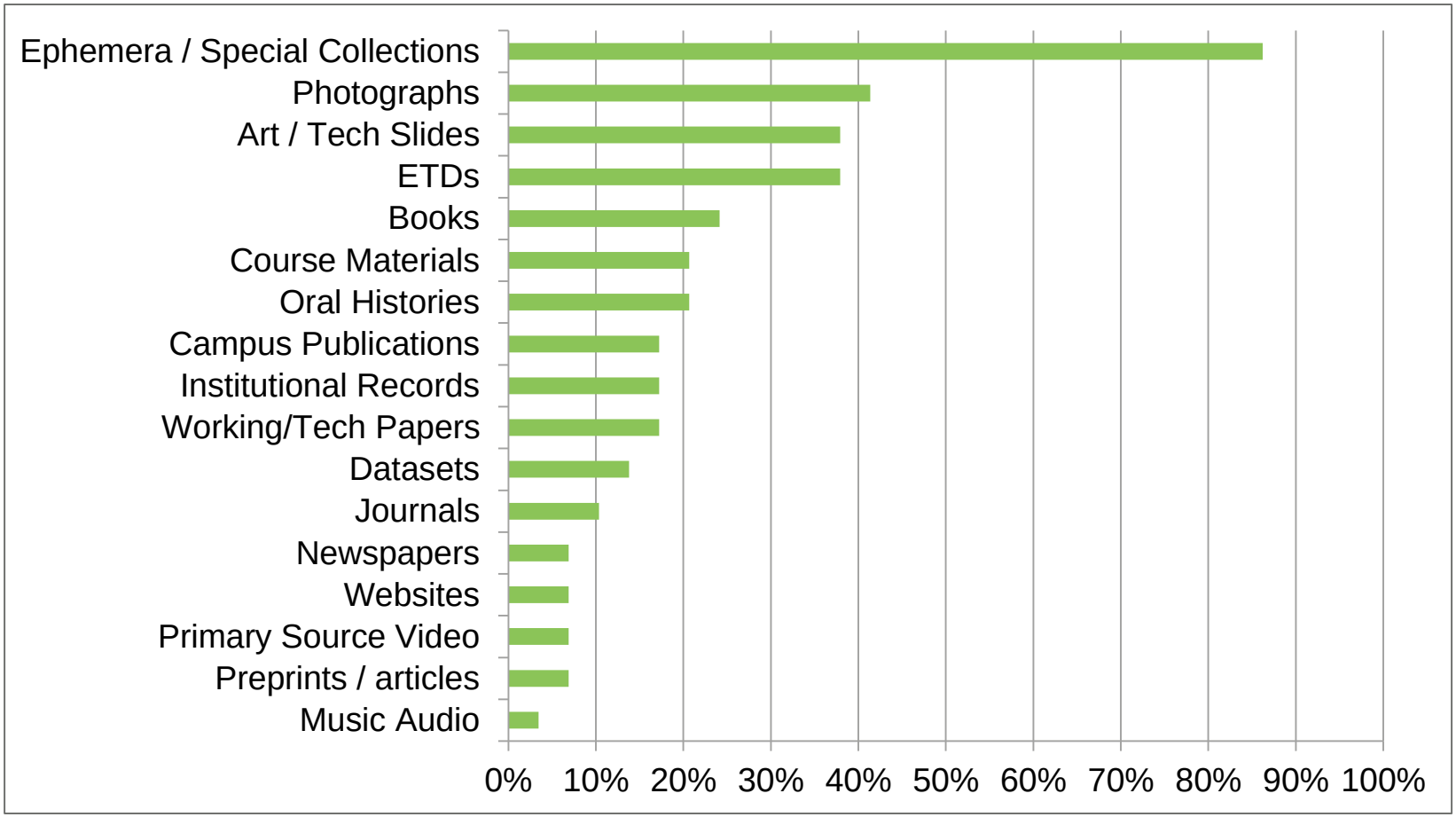
Digital preservation is the series of management policies and activities necessary to ensure the enduring usability, authenticity, discoverability, and accessibility of content over the very long-term. The key goals of digital preservation include:

Usability	Authenticity	Discoverability	Accessibility
• the intellectual content of the item must remain usable via the delivery mechanism of current technology	• the provenance of the content must be proven and the content an authentic replica of the original	• the content must have logical bibliographic metadata so that it can be found by end users through time	• the content must be available for use to the appropriate community

FULL MANAGED DIGITAL PRESERVATION

ANALYSIS FROM THE STUDIES

Survey Analysis



DIGITAL CONTENT CONSIDERED IN STUDIES

Survey Analysis

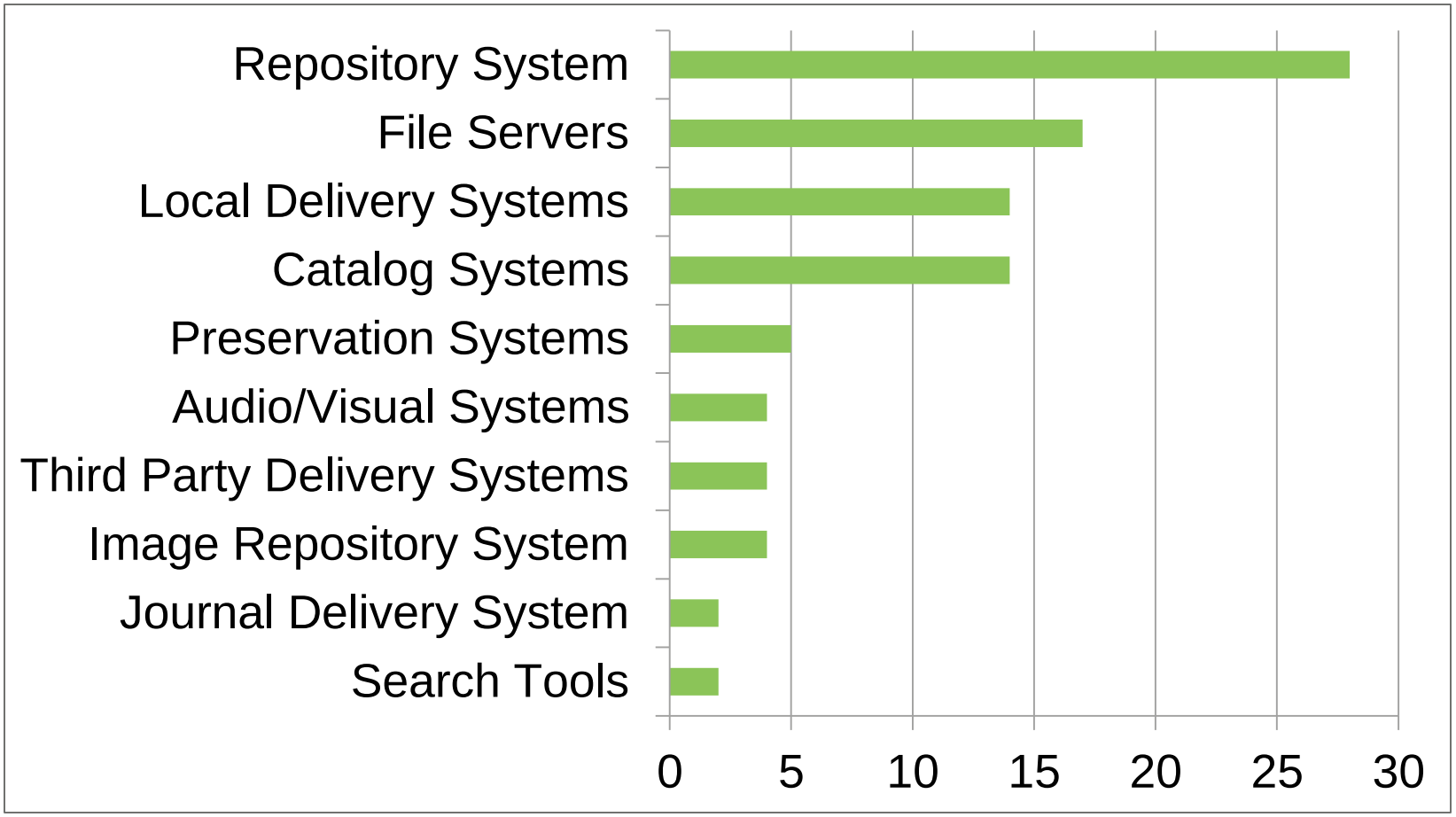
The 27 institutions, and at least 30 projects, reviewed were using:

36 distinct pieces of software

94 instances of software

3.13 pieces of content management software, on average

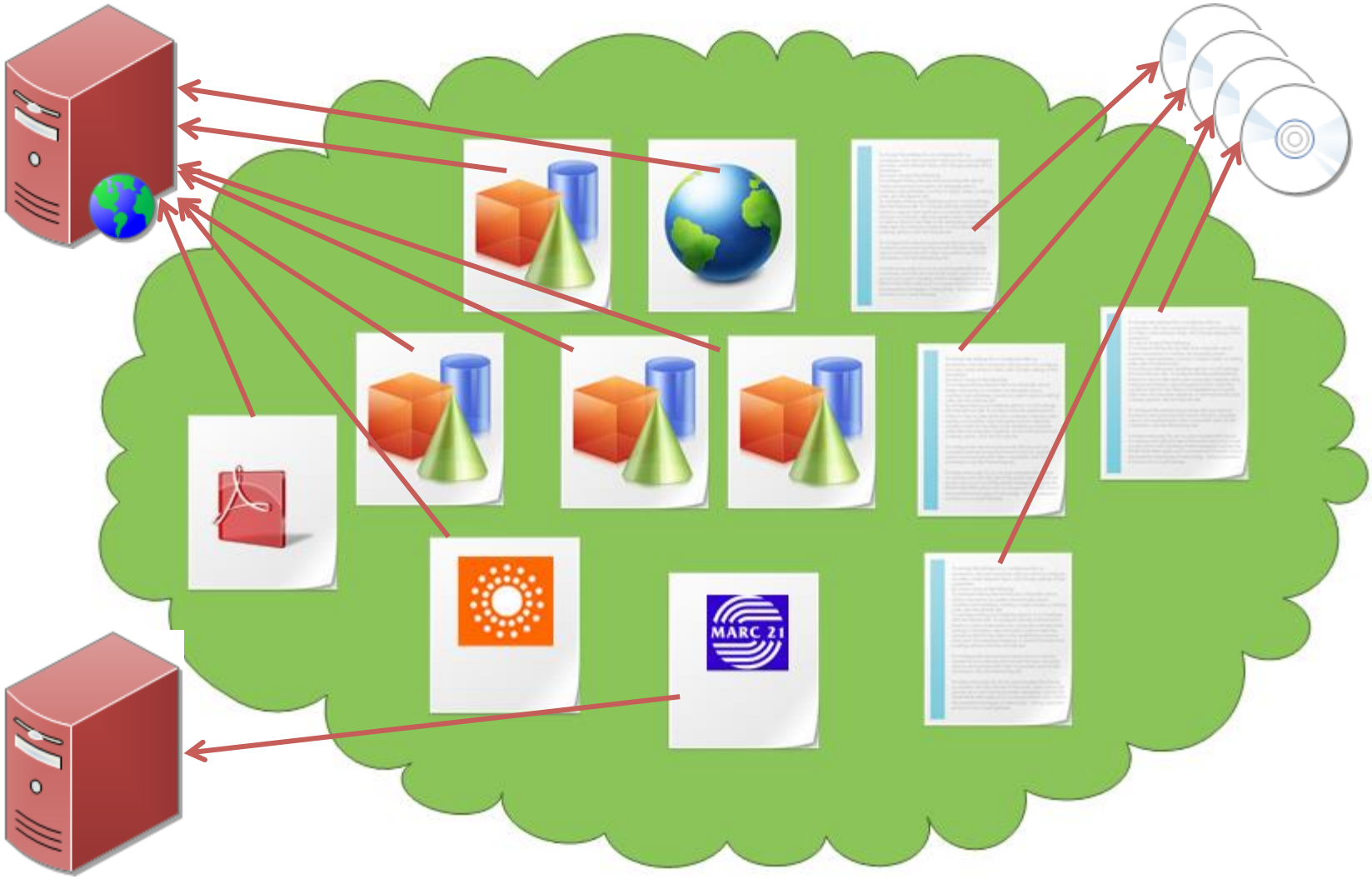
Survey Analysis



COUNT OF INDIVIDUAL INSTANCES OF TYPES OF SYSTEMS IN USE ACROSS ALL ORGANIZATIONS

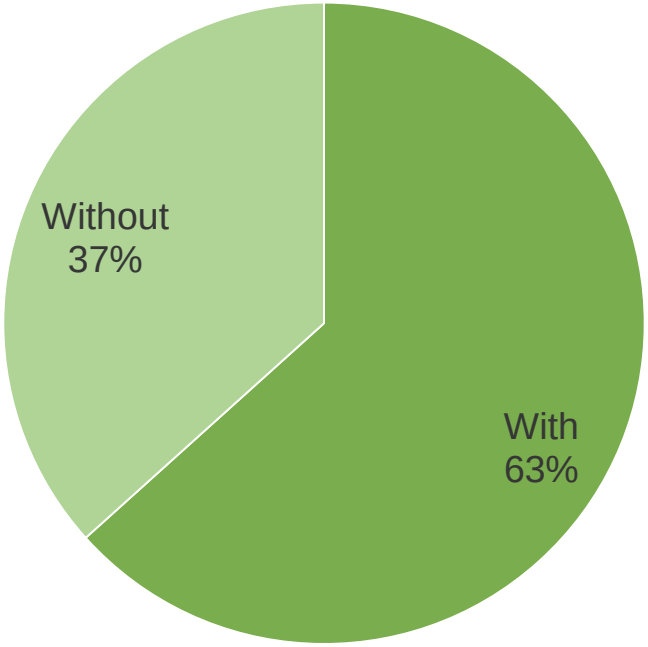
Survey Analysis

Survey Analysis

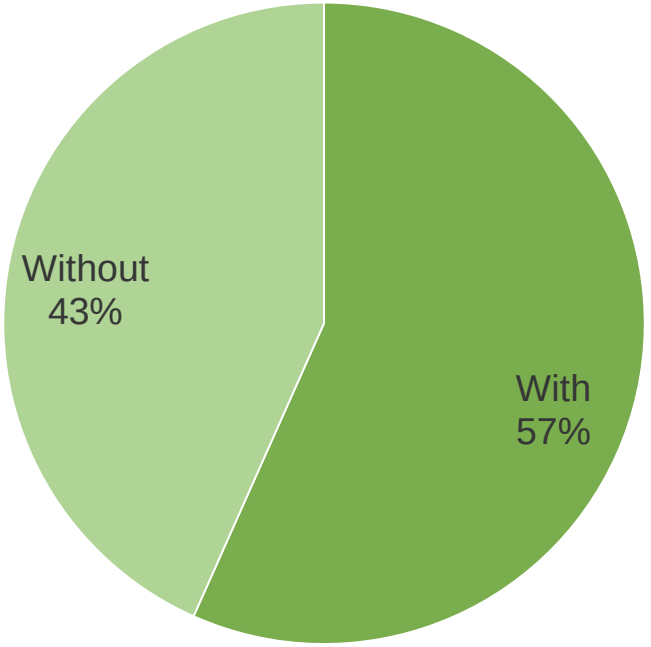


Survey Analysis

Institutions with InterInstitutional Collaborations

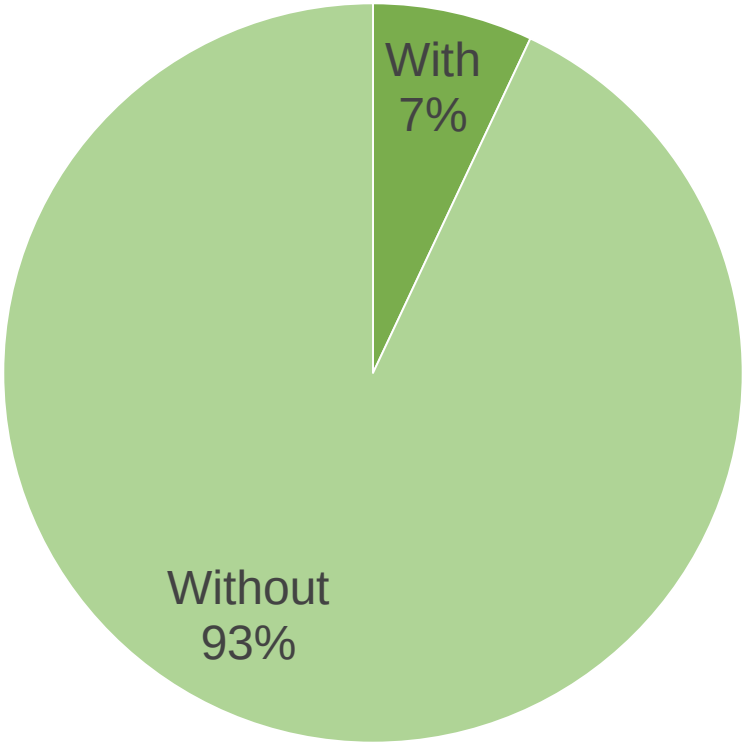


Institutions with IntraInstitutional Collaborations



COLLABORATION

Survey Analysis



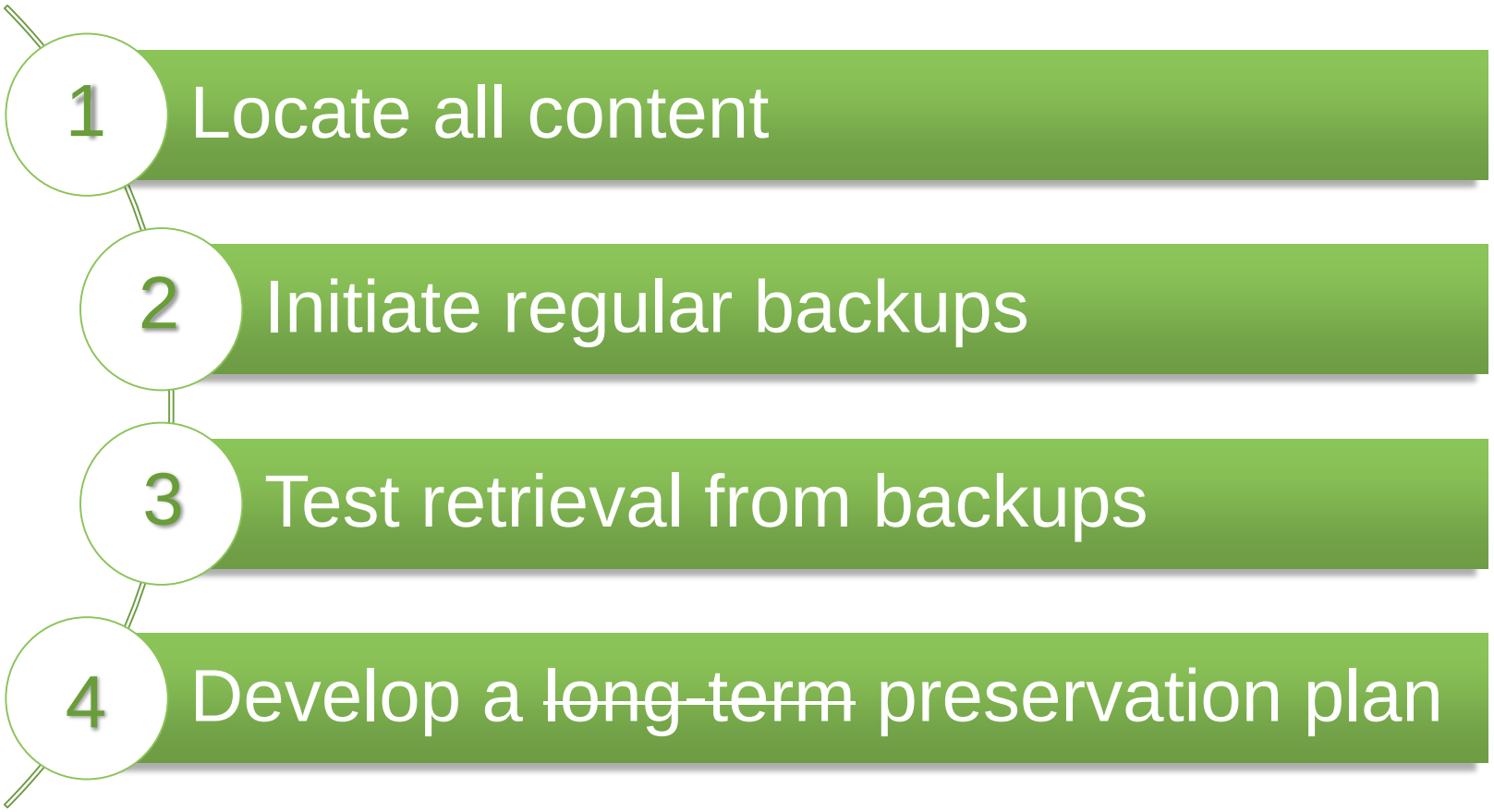
INSTITUTIONS WITH AND WITHOUT PRESERVATION POLICIES

Survey Analysis – Additional Themes

- The size of a cultural heritage organization is neither a predictor for risk nor amount or quality of digital content.
- Many cultural heritage organizations do not have staff to support either preservation or access systems in-house.
- Cultural heritage organizations would benefit from a turn-key solution that provides both access and preservation for a large variety of formats and content types.
- Cultural heritage organizations do not often have a sustainability plan associated with their digital content. The digital content is not considered a product that must be sustained, it is considered another outlet for their special collections.

INITIAL STEPS TO PROTECT CONTENT

Initial Steps to Protect Content



Initial Steps to Protect Content

PRESERVATION (OR PROTECTION) PLANNING

1. Who: Identify the key players involved with long-term preservation of the targeted content
2. What: Describe or characterize the collection and content
3. Where: Document the locations of all the copies of the content and metadata.
4. When: Document the targeted preservation timeframe and impact of loss.
5. How: Document how the key content management and preservation tasks will occur.

Survey Analysis

Initial Steps to Protect Content

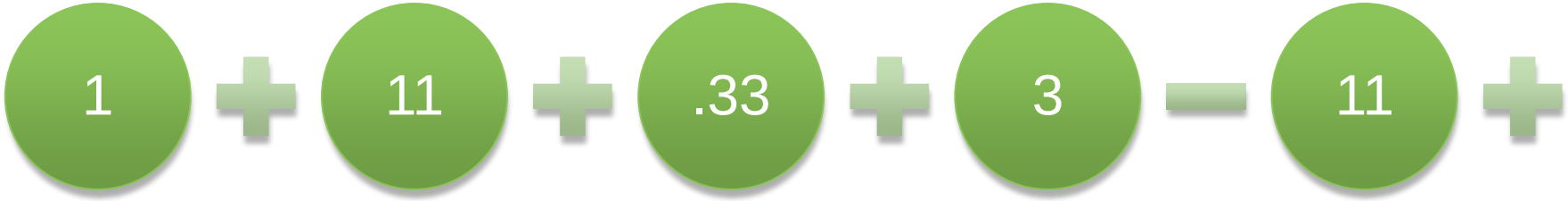
WHY BACKUP?

Initial Steps to Protect Content

CALCULATING A BACKUP SOLUTION

Initial Steps to Protect Content

CALCULATING A BACKUP SOLUTION

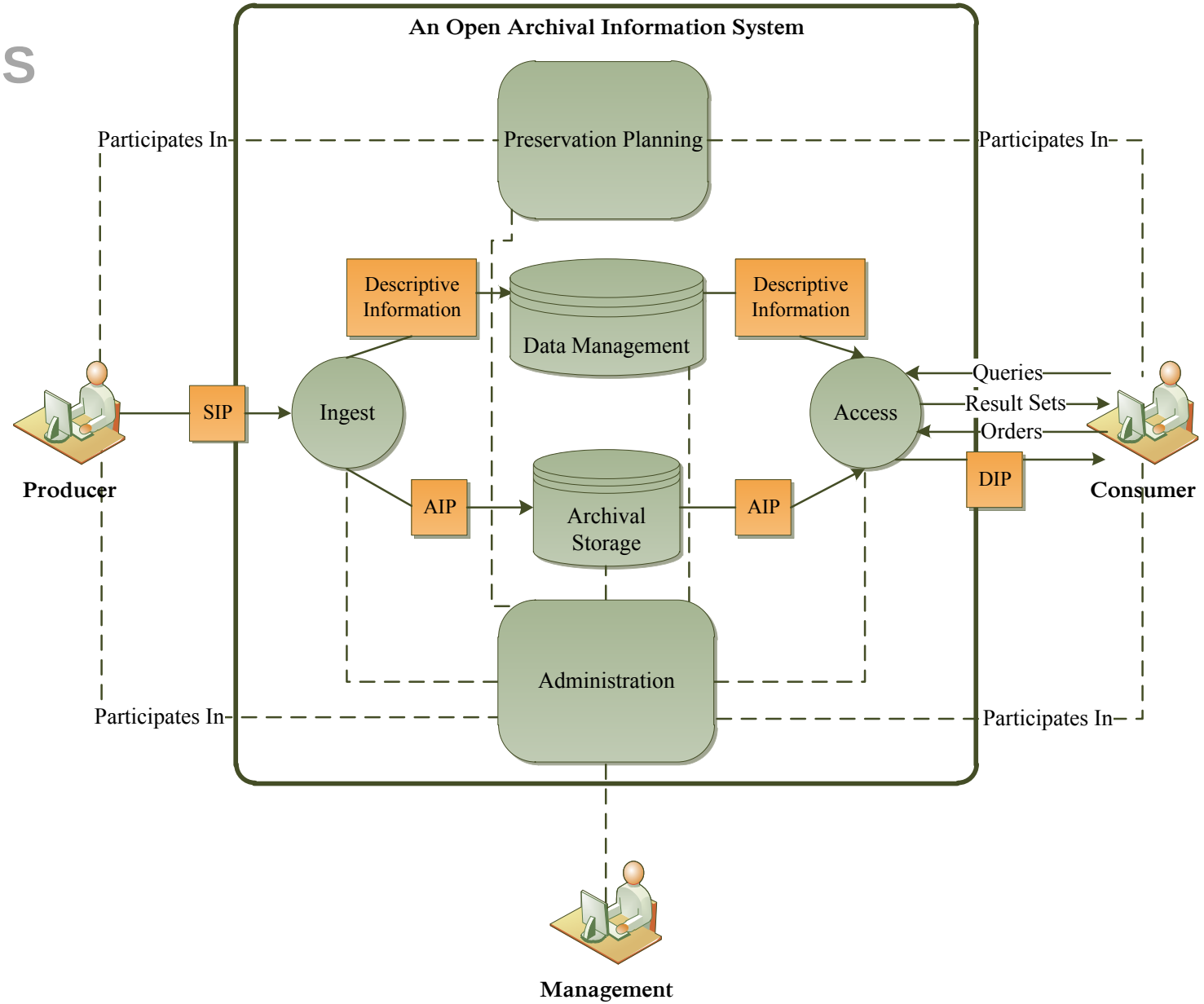


PRESERVATION

A Preservation Model

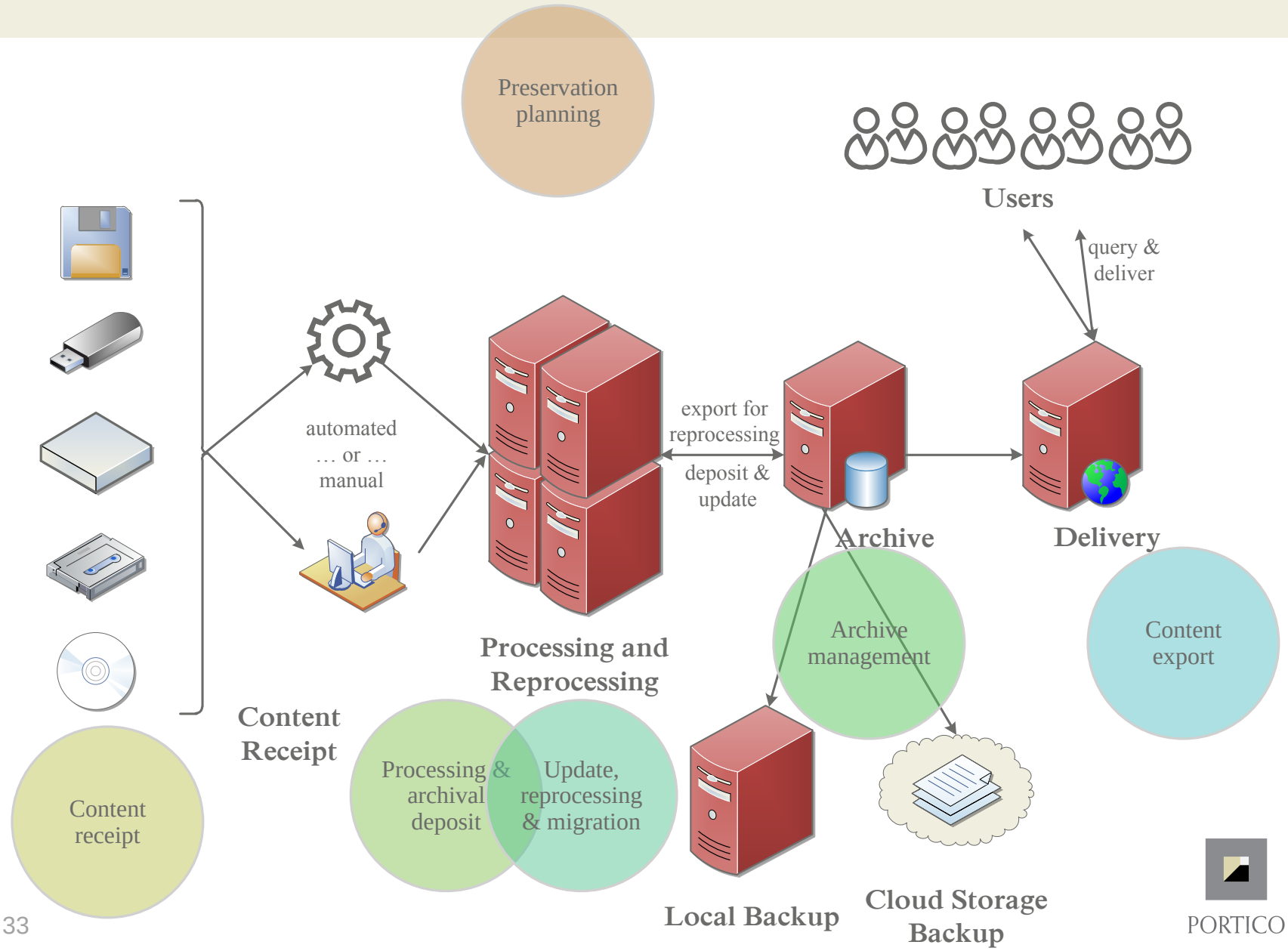
ITHAKA
JSTOR | PORTICO | ITHAKA S+R

OAIS



PORTICO

A Preservation Model



PORTICO

A Preservation Model for Cultural Heritage Organizations

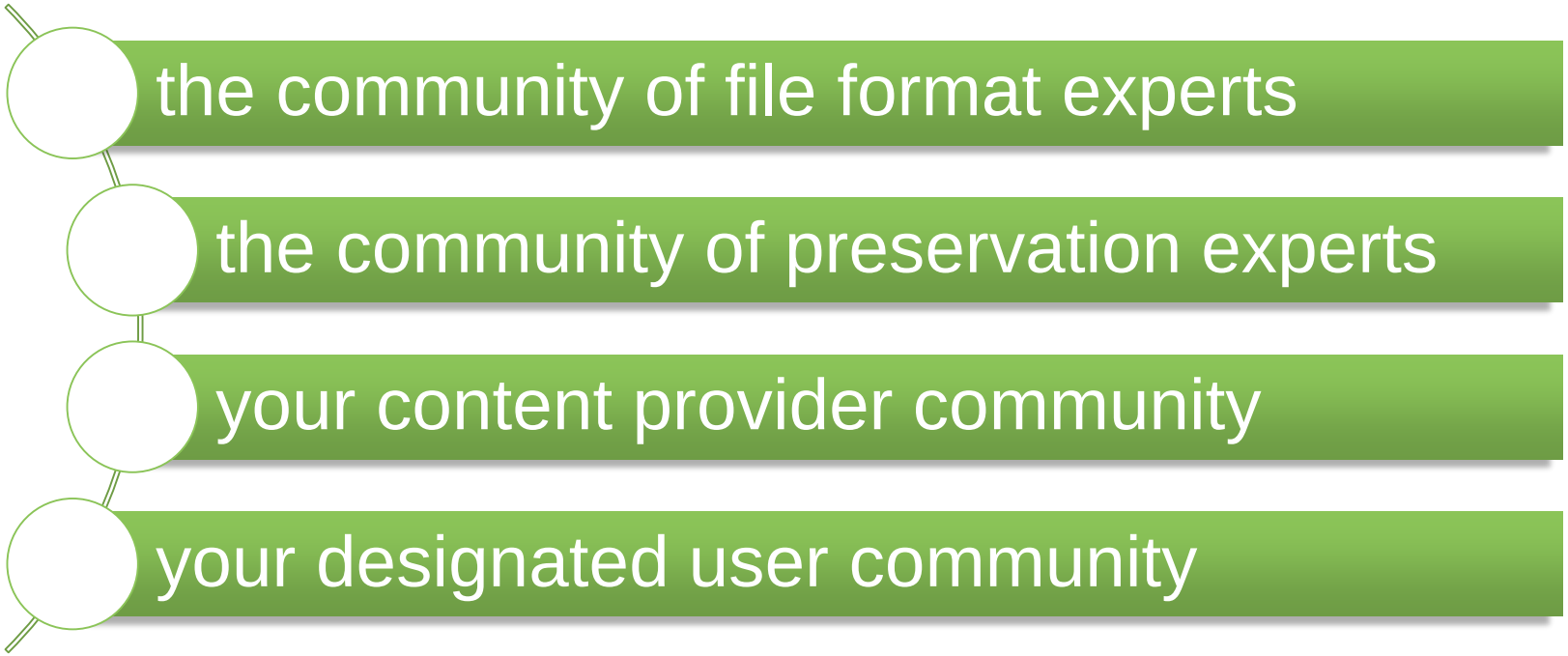
A Preservation Model

SELECTING SOFTWARE

- Will the software meet your input throughput needs? Will it meet your batch or otherwise loading needs?
- Will the software meet your output throughput needs?
- How complicated is the software to manage? Do you have appropriate staff to both install the software and maintain it over time?
- Will the software capture the preservation metadata you have identified as necessary in your policies?
- Can the software support maintaining the original master versions of your content files and the web-ready versions of your content files side-by-side with the metadata for the files?
- Can the software export the original master version of your content files with the metadata for those files?
- How much does the software cost initially? Consider both internal costs such as staff time and external costs.
- How much will it cost to maintain? Consider both internal costs such as staff time and external costs such as licensing fees.

A Preservation Model

COMMUNITY MONITORING



A Preservation Model

limit

limit the types of formats allowed in the collection to those with a proven long-life

migrate

migrate files in troublesome formats to new formats

both

accept all file formats, but through written policies, assign different preservation commitments to different types of files

CONTENT FORMATS

A Preservation Model

ITHAKA
JSTOR | PORTICO | ITHAKA S+R

PRESERVATION POLICIES

<Preservation Policy Title>

1. **Policy Statement**

   1.1. <Paragraph 1 in high level policy statement>

   1.2. <Paragraph 2 in high level policy statement. This section should be short, but accurate and to the point. It should provide guidance to operations staff as they do their jobs.>

2. **Implementation Examples**

   2.1. <Example 1>

   2.2. <Example 2 – examples should describe practical decisions made based upon this policy.>

3. **Document History**

   3.1. Approved by: <Name of approver>

   3.2. Last Review Date: <Date policy was last approved>

   3.3. Reviewed by: <Names of people who reviewed the document at ita last review date>

   3.4. Change history:

Version	Date	Change	Author
<version #>	<date finalized>	<Highlight the changes made to the document>	<author of the changes>
<version 2>*	<date finalized>	<Highlight the changes made to the document>	<author of the changes>

* An approved version of this document.

PORTICO

A Preservation Model

ITHAKA
JSTOR | PORTICO | ITHAKA S+R

METADATA & PACKAGING

PORTICO

A Preservation Model

METADATA &
PACKAGING

PACKAGING

A Preservation Model

METADATA & PACKAGING

PACKAGING & PRESERVATION METADATA

- The intellectual unit represented by this metadata file is a digitized book.
- It was scanned by Joe on this date.
- It was ingested into the repository on this other date.
- Jane Smith granted us preservation rights to it on this other date.

...

Preservation and Packaging Metadata File

- These TIF files are page images.
- The TIF file named XYZ is page 1. It is a valid TIF and has a checksum of 123456.
- The TIF file named ABC is page 2. It is not a valid TIF and has a checksum of 78910.

...

- These JPG files are figures.
- The JPG file named MNO is the 2nd figure on page 2. It is a valid JPG and has a checksum of 234567.

...

- This PDF file contains page images.
- The page images are built from TIF files XYZ, ABC, etc. and JPG figure graphics MNO, etc.

...

- This MARC file is the bibliographic record for the book.

...

- This XML file contains the full-text of the book.
- It uses the QRS DTD.
- It is named JKL and has a checksum of 555555.

...

ORTICO

A Preservation Model

PACKAGING

- METS
- DIDL
- OAI-ORE

PRESERVATION METADATA

- PREMIS

A Preservation Model

- Access rights are not preservation rights
- Get preservation rights at the point of deposit
- Document preservation rights in the preservation metadata
- Tricky area – embedded content
- Tricky area – research and privacy rights

RIGHTS

A Preservation Model

COSTS

A Preservation Model

RECOVERING COSTS

- Allowing sponsorship of the collection
- Permitting advertising on the collection
- Encouraging donations to support the collection
- Building an endowment
- Creating premium services for purchase, the revenues from which can subsidize the preservation service.
- Enlisting support from funding bodies, scholarly and professional societies, and publishers
- Charging for access to the collection and using the revenue to subsidize the preservation costs.
- Charging for participation in the preservation service
- Relying upon support from a parent organization or government

A Preservation Model

- Do-It-Yourself
- Collaboration
- Third Party Preservation Service

ORGANIZATIONAL MODEL

A Preservation Model for Cultural Heritage Organizations

PORTICO

THANK YOU.

Amy Kirchhoff
amy.kirchhoff@ithaka.org

http://www.portico.org