



---

## INTRODUCTION

This is a handout paired with the ALCTS Preservation Week Webinar, Protecting Future Access Now—Models for Preserving Locally Created Content by Amy Kirchoff ([amy.kirchoff@ithaka.org](mailto:amy.kirchoff@ithaka.org)).

---

## RESEARCH

Webinar Slides & Handouts: <http://www.portico.org/alcts-webinar/>

White Paper supporting the presentation:

“Preservation of Digitized Books and Other Digital Content Held by Cultural Heritage Organizations.” A report for the NEH and IMLS resulting from a grant from the “Advancing Knowledge: The IMLS/NEH Digital Partnership” given to Portico and Cornell University Library. (Mar 2010). <http://www.portico.org/digital-preservation/presentations-publications/>

JISC Digitization Programme Preservation Study: <http://www.dpconline.org/advocacy/knowledge-base/594-digitisation-programme-digital-preservation-study>

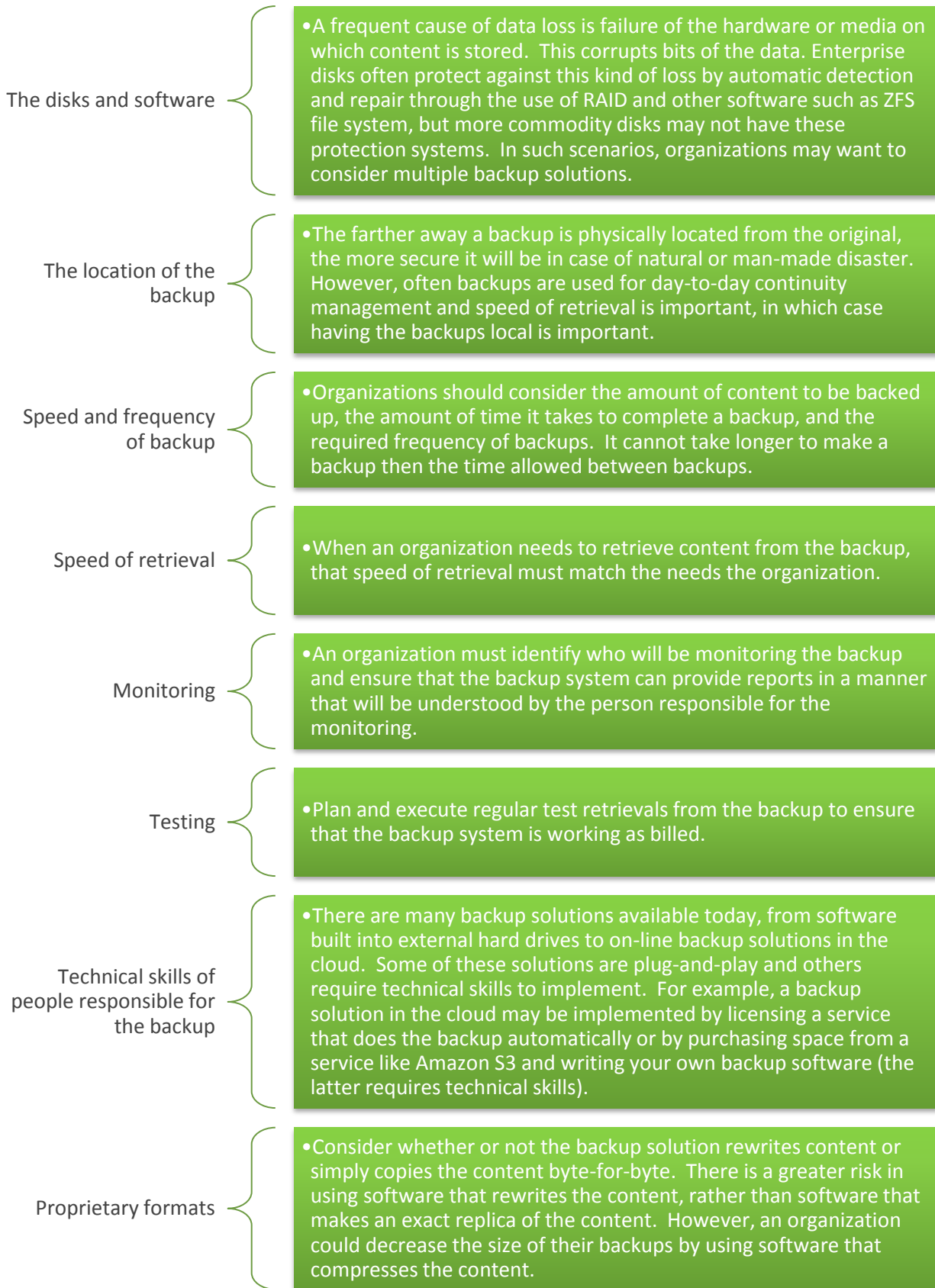
---

## IMPLEMENTING BACKUP AND BYTE-REPLICATION

Backup and byte replication are well-understood solutions. Many cultural heritage organizations may be able to get robust backup from their parent institution. In general, backups provide solutions to two problems:

1. **User error recovery** – a user or system accidentally deletes or modifies some files and those few files need to be copied out of the backup and back onto the system. In this regard, currency of content is very important. The backup must have current versions of the files, or it cannot serve this purpose well. In order to support this type of file-by-file retrieval of current files and to quickly make the backups, most backup solutions implement a type of delta backup, such that only items that have been changed since the last backup are copied. The organization should expect retrieval of those few files to be relatively fast for the backup to effectively meet this need.
2. **Disaster recovery** – a natural or man-made disaster destroys the original copies of the content and the system needs to be entirely rebuilt. In this regard, currency is less important, as the organization will be spending considerable time rebuilding the system (perhaps even the machine room) and loss of a week or two of updates to the content will not significantly impact time to recovery.

Several things to consider when selecting a backup solution are:





## PRE-PRESERVATION ANALYSIS & PLANNING

As a result of this analysis, institutions will be able to make informed decisions about the length of time the collection must be protected and therefore the amount of investment to be made in that protection (for example, is backup and/or byte replication sufficient, or does the collection need long-term, managed digital preservation.) The report created through answering the following questions can also be used to form the basis of a preservation policy for the content.

### WHO: IDENTIFY THE KEY PLAYERS INVOLVED WITH LONG-TERM PRESERVATION OF THE TARGETED CONTENT

Our surveys have shown that, especially when multiple partners are involved in managing content, there is an opportunity for misunderstandings to arise around which party is responsible for which element of the content. Often the role responsible for managing the files is different from the role responsible for managing the intellectual content of the collection. Therefore, for each digital collection, it is important to identify the key players involved with the development and long-term management of the content.

1. Who is writing the policy and plan?
2. Who has responsibility for maintaining the intellectual content of this collection (e.g. making corrections to metadata or content files)? Who has curation responsibilities and is the advocate for the collection?
3. Who has responsibility for maintaining the bytes of the files in this collection (e.g. identifying and fixing corrupted files)?
4. Who approved this policy and plan?
5. Who will use the content in the short and long-term?

### WHAT: DESCRIBE OR CHARACTERIZE THE COLLECTION AND CONTENT

Per the definition of digital preservation, being able to trace the authenticity of an object in the collection is important. From a practical point-of-view, this provides information to those people who will be managing the content in the future, but may not have been involved in its original creation. The ability to quickly characterize a collection is also very important when it becomes necessary to consider all digital collections at one organization and organization-wide preservation solutions.

6. What is the content and from where did the content originate?
7. What file formats, including metadata formats, are present?
8. How many items are in the collection? How large is the collection on disk?



**WHERE: DOCUMENT THE LOCATIONS OF ALL THE COPIES OF THE CONTENT & METADATA.**

Our surveys have found that there are often many versions of content “around.” In order to manage all this content in a sensible way, it is important to identify where all the content is and the purpose of the copy at each location.

- 9. Where are the high resolution master copies of the descriptive metadata kept?
- 10. Where are the master copies of the content files kept?
- 11. Where are all the copies of the content, including backups, and how are the copies of the content related?

**WHEN: DOCUMENT THE TARGETED PRESERVATION TIMEFRAME AND IMPACT OF LOSS.**

Not all content must be preserved forever, some content can be protected for a limited time, after which its status will be re-evaluated. Identifying what might happen if the content were irretrievably lost will help answer the question of how long it must remain available. Other factors include user demand and organizational mission.

- 12. How long should the content be available for use?
- 13. If the content is irretrievably lost, what are the repercussions?

**HOW: DOCUMENT HOW THE KEY CONTENT MANAGEMENT & PRESERVATION TASKS WILL OCCUR.**

It is important to make thoughtful decisions about how to manage the collection. A closed collection may be deposited into a read-only archive, whereas an open collection that will have updates made to it must be preserved in an archive that allows updates. Having all parties responsible for the content answer this set of questions together will ensure that everyone agrees on how the content will be managed.

- 14. How will the collection be created (perhaps draw a diagram of the workflow)?
- 15. How will the collection be maintained (perhaps draw a diagram of the workflow)?
- 16. Do you expect the content files to be migrated in the future?
- 17. May the content files be deleted? Added to? Updated?
- 18. May the descriptive metadata be deleted? Added to? Updated?
- 19. How will you track who did what and when to the content, if this is important to your organization?
- 20. How do you associate the master copy of the descriptive metadata with the high resolution copy of the content files and how will you move these two items around together?



## RECOMMENDED PRESERVATION POLICIES

Required metadata

Content selection

Securing preservation rights

Modification and deletion of preserved content policy

Collection specific preservation policies

Provider initiated update policy

Designated community and feedback policy

Documentation and policy review cycle

Migration and emulation policies

Hardware and software lifecycle and refreshment policy

Identifier usage policy

Problem resolution escalation path

Public disclosure of agreements policy

Software development and content processing quality control policies

Replication and backup policies

Roles and responsibilities

Succession or end of life policies



# POLICY TEMPLATE

<Preservation Policy Title>, v. <version #>

## <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 its 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.

Last update: <date last approved>

Page 1 of 1

## SYSTEMS IN USE ACROSS ALL INSTITUTIONS SURVEYED

Type of Software System	Specific System	Instances
Repository Software	CONTENTdm (local, hosted, pro)	12
Repository Software	Fedora	6
Repository Software	DSpace	6
Repository Software	ExLibris DigiTool	1
Repository Software	Innovative's Symposia	1
Repository Software	BePress Digital Commns	1
Repository Software	VITAL	1
<b>Repository Software Total</b>		<b>28</b>
Image Repository	MDID	2
Image Repository	Luna	1
Image Repository	Artesia	1
<b>Image Repository Total</b>		<b>4</b>
Search Tools	Solr	1
Search Tools	DTSearch	1
<b>Search Tools Total</b>		<b>2</b>
Delivery	ETHOS (delivery)	1
Delivery	Bespoke Delivery	8
Delivery	Drupal	3
Delivery	Static Web pages	2
<b>Delivery Total</b>		<b>14</b>
Journal Delivery	OJS	2
<b>Journal Delivery Total</b>		<b>2</b>
Preservation	Bespoke Preservation	4
Preservation	Quantum Digital Archive	1
<b>Preservation Total</b>		<b>5</b>

Type of Software System	Specific System	Instances
3rd Party Delivery	JSTOR	2
3rd Party Delivery	Cengage	1
3rd Party Delivery	ProQuest	1
<b>3rd Party Delivery Total</b>		<b>4</b>
A/V	iTunesU	2
A/V	Streaming Server	2
<b>A/V Total</b>		<b>4</b>
File Server	File Server	17
<b>File Server Total</b>		<b>17</b>
Catalogs	IRIS (MD in FMPro)	1
Catalogs	CALM	1
Catalogs	MODES Catalogue	2
Catalogs	Catalogue -- Unknown	2
Catalogs	Extensis Portfolio	2
Catalogs	Relational DB	1
Catalogs	Allegro DB	1
Catalogs	OPAC	1
Catalogs	Tec-Rec	1
Catalogs	SIFT	1
Catalogs	MINISIS	1
<b>Catalogs Total</b>		<b>14</b>
<b>Grand Total</b>		<b>94</b>