

Digital Preservation with **libsafe**

libnova – July 2014

Paseo de la Castellana, 153
28046 – Madrid

Tel: 91 449 08 94
Fax: 91 141 21 21
info@libnova.es

The logo for libnova, featuring the word "libnova" in a lowercase, sans-serif font. Above the letter "o" is a stylized icon of an open book, represented by two curved lines forming the pages and a horizontal line for the spine.

Digital preservation with libsafe

This document is

CONFIDENTIAL / AUTHORIZED USE ONLY

and should not be reproduced or disclosed without prior written consent of **libnova**, and in any case excluding considerations of purpose and scope of the document itself.

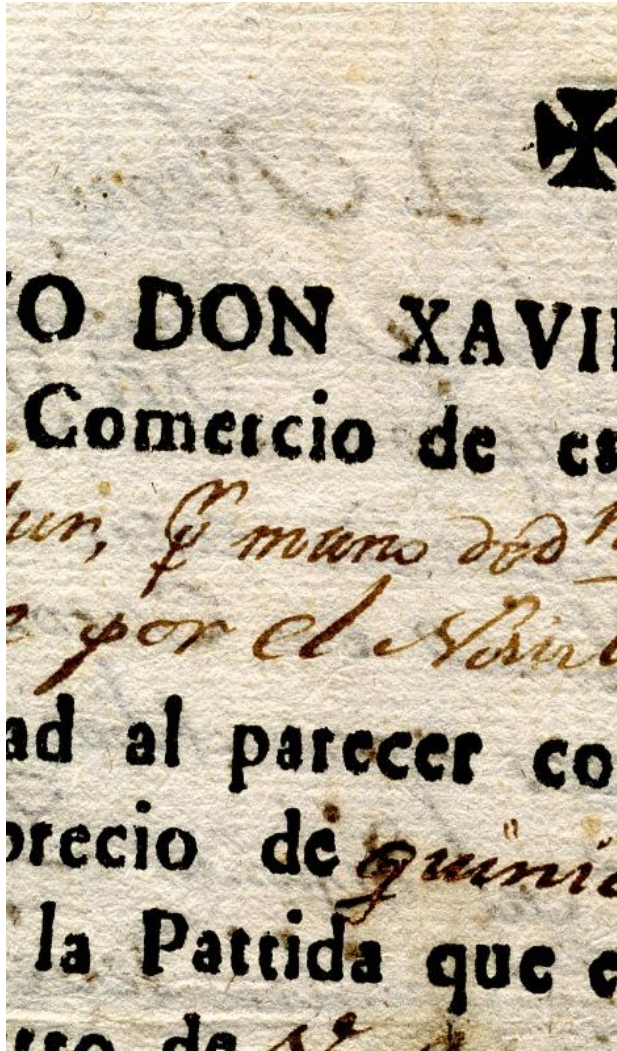
This document and its attachments contain confidential or legally privileged information and is intended only to authorized personnel under NDA.

You are not allowed to read or hold a copy if you receive it in other case.

Additionally, in no event may you modify, distribute, copy or disclose its content except as provided above.

The images contained in this presentation are owned or licensed by **libnova**, or have been released to the public domain for reuse.

Preserving digital objects: a real case



2007

The historical collection is massively digitized.
The resulting masters are stored in CD/DVD and HDs.

2014

20% of the storage media are already degraded.
Some of the used formats do not longer exist.
The central catalog, the master mediums and the
objects stored in each one are not related any more.

It's only been 7 years

How many of these masters are still usable?
Now we have the methods and the technology to
prevent this from happening again.

Content

1.

Why are Masters different ?

Masters originated during digitization processes have special features that affect the way they have to be managed and stored.

2.

Traditional methods are not valid any more.

Traditional methods for managing and storing digital information are very costly and ineffective when the objective is to handling and preserving masters for the long term.

3.

Digital Preservation is the solution.

In order to solve this complexity, specific methodologies and rules have been developed. They are very effective but unfortunately their implementation is very complex.

4.

libsafe is preservation made easy.

libsafe implements a digital preservation model based on OAIS and ISO 14.721 in a complete, simple and sustainable way.



Masters main characteristics

- **They need a lot of storage capacity**

A single book may require tens of Gigas. Much more space than the one needed for derivatives for dissemination and other digital data.

- **Diversity of formats**

Both in the storage media and in the objects content.

- **Low frequency of queries**

The potential problems, both for accessing and formatting, will most probably be discovered when is too late to solve them.

- **The storage media are kept off-line**

Hence cataloging, documenting and including metadata are extremely difficult tasks.

- **They have big value**

Both for the preservation of the physical object and to avoid having to pay again the digitization costs.

How can masters
be managed to
guarantee their
preservation and
future usability ?



How have they been traditionally managed ?



Storing them in backup tapes, DVDs or external offline disks.

- Dozens of different storage media types, file formats and compression methods are available in the market.
- With an average lifespan of 5-7 years.
- Bound to obsolescence and degradation.
- If corruption happens, it is then passed to all the copies.
- Cataloging, documenting, finding, retrieving and auditing the content are extremely difficult tasks.



How have they been traditionally managed ?



Storage in disk arrays and servers

- High cost: conceived to be used in production environments with quick and frequent access.
- Files can be easily deleted or modified.
- Very complex backups, even sometime impossible due to the high volume of information.
- If corruption happens the backup is also corrupted.
- Obsolescence and lack of metadata risks, are not mitigated.



How have they been traditionally managed ?



Tailor made projects

- Very complex projects.
- High human resources consumption, both in the implementation and maintenance phases.
- High cost and low sustainability.
- Experience from best practices and other customers know-how is not included.
- Usually only partial preservation processes and solutions are applied, therefore the whole Master lifespan is not considered.



La solución: Preservación Digital

In order to guarantee the future use* of digital assets
specific methodologies, technology and activities
are required

This is **digital preservation**

(*) NOTE: For the future use of a digital asset to be possible it is required:
availability, integrity, safety, authenticity, and accessibility,
and capability to represent and view its content.

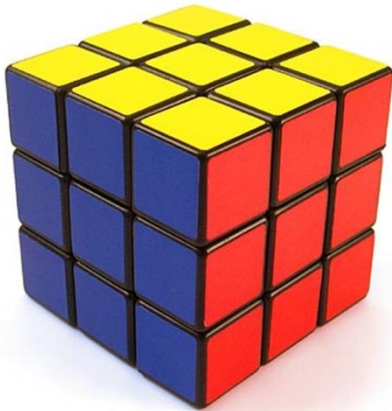


Digital preservation minimizes risks

Reliability regarding ...		backup	CD/DVD/HD	Storage	Preservation
Storage related risks	Storage media degradation	XX	XX	—	✓
	Product problem	X	X	X	✓
	Risk of loss of the physical media	—	—	—	✓
	Media obsolescence	XX	X	—	✓
Files related risks	Data corruption	XX	X	—	✓
	Accidental modification/deletion	✓	✓	X	✓
	Format obsolescence	X	X	X	✓
	Need of format migration	X	X	X	✓
Object related risks	Defective original objects	X	X	X	✓
	Lack of, or bad metadata	X	X	X	✓
	History of changes in the data	X	X	X	✓
	Cataloging and finding objects	X	X	—	✓
	Access safety and audit	X	X	—	✓



Digital preservation has many facets



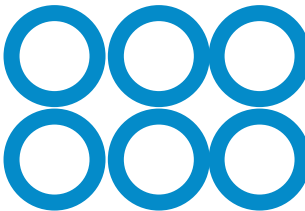
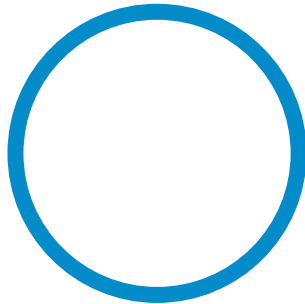
- **Huge volume of non-structured information**
Hundreds of thousands of images, audio files, and scanned documents, that occupy hundreds of Terabytes.
- **Managing multiple copies**
Keeping them independent, certified and audited.
- **Keeping control on the collection**
Periodic and complete information about the content of the collection. Capability to adapt the preservation plan to new methods and processes.
- **Checking the material validity at entry point**
Viruses, authorizations, names and any other aspects that can be relevant for the future.
- **Evolving objects and formats**
Approximately every 7 years the Industry change formats and storage mediums, but the content must remain accessible.

**Preservation is a complex process
that goes far beyond simple storage and archiving**



libnova: preservation made easy

libnova has developed a **digital preservation** platform



Simple

Because it has been developed from a real world project. We adjust the standards to the real daily processes needed to manage a collection of digital masters.

Complete

Because it has an end to end approach in a manner consistent with OAIS and ISO 14.721, from quality control to audit and file transformation.

Sustainable

libnova's experience and technology surveillance are reused and enhanced in every new project. All repetitive and/or complex processes are automated to make them more efficient and safe.



User friendliness, safety for your digital collection and peace of mind, all in one software.

INGESTION AND
DISSEMINATION

AUDIT AND
AUTOHEALING

TRANSFORMATION
AND EVOLUTION

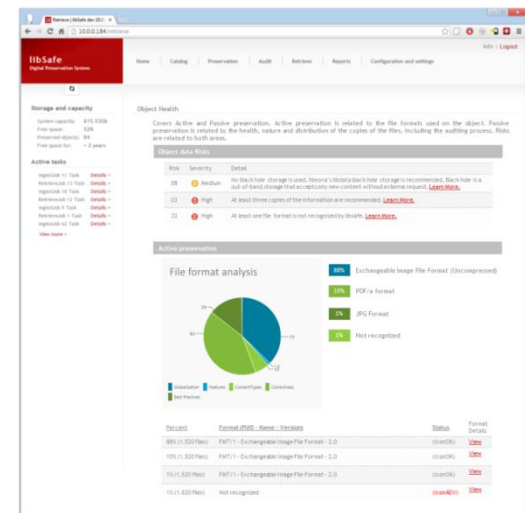
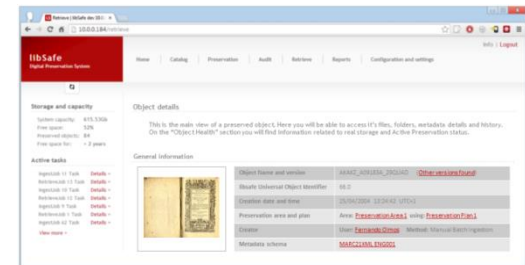
CATALOG AND
PREVIEW

- Checks that all the information is correct according to your preservation plan..
- Executes your dissemination policy in different copies and storage systems..

- Continuously verifies that all the copies of the stored information are identical and equal to the original ones..
- Detected risks are reported immediately and solved when it is possible.

- Technology surveillance for optimal metadata , formats and preservation processes..
- Aided digital migration when the obsolescence point is reached.

- An integrated catalog allows easy searching, previewing and retrieving all the preserved objects, hence guaranteeing all the contents safety.





The storage array which is a perfect matching with **libsafe** for your digital preservation system

- **Specifically designed for digital preservation**

Optimal access, redundancy and safety features.

- **Fully integrated**

Both with **libsafe** and with your data center infrastructure.

- **No useless features**

Useless or even preservation counterproductive features have been eliminated (RAID5, compression, de-duplication).

- **Want more safety ? Hybrid systems**

libnova recommends hybrid dissemination architectures based on **libdata** and your current storage solution providers.

- **And, of course, at the best price**

Thanks to its simplicity, you get the highest reliability at the best price.





A **successful** preservation project

- **Where should I start ?**

From the simple to the complex

In preservation, experience is very important, and doing anything is always better than doing nothing. Start with a collection that has homogeneous formats and objects, so that you can concentrate in adapting your organization to the preservation approach. Once your organization is fully aligned you can treat heterogeneous objects in a much more efficient way.

- **What are the first actions to have a successful project ?**

Selection of material, formats and metadata

Selection of the material to be preserved (if you cannot preserve everything at once), selection of formats (choosing the most standards, so that they have more longevity) and including metadata. In this way you can have your master's collection under control and leave for a later stage the most complex material .

- **How can I guarantee my masters for ever ?**

Digital collection: control and evolution

In a fast evolving digital world, it is difficult to have a technology that lasts for ever. In order to guarantee the life of your information beyond technology obsolescence there are two main steps to be taken. The first one is to keep your collection properly controlled and documented. The second one is to take the correct actions so that you can guarantee the integrity of your collection until the next technological change happens. And then you can make the appropriate decisions to guarantee success. Libsafe is the perfect tool to help you in this process.



Libsafe preservation platform: licencing

- **libsafe**
 - License for use + Volume (Tb) of preserved objects
 - Related to the size of preserved masters, regardless the number of copies.
 - The base license includes 5 Tb.
- **Libdata**
 - Related to the total volume of requested storage
 - Systems for 36 and 72 disks (min. 12) of 4Tb
 - Gross capacity from 48 Tb to 288Tb for each unit
 - Unlimited number of units within the same storage pool
 - High density; up to 2.8Pb per rack.



libsafe preservation platform: features

Ingestion processes	Sanitization of materials	<p><i>Sanitization</i> sets formal aspects of the material to be ingesting.</p> <ul style="list-style-type: none">• Verification and correction of file permissions• Verification of illegal characters in file names and folders• Verification of the maximum size of folder paths• Deletion of system files and temporary application files and folders• Inventory of file formats with DROID• Extensible with user-defined controls for specific materials
	Checks in ingestion phase	<p>The ingestion checks verify the validity of the content to be ingesting:</p> <ul style="list-style-type: none">• Checks at object, file or folder level• Existence checking and verification of valid size ranges• Name and character convention checking according to preservation plan• Format and content validity check with JHOVE• Extensible with user-defined controls for specific materials
	Metadata	<ul style="list-style-type: none">• Preloaded with Dublin Core, Marc21 and ISAD (G) standard schemas• Ability to include custom metadata schemas defined by the user• Ability to read custom XML files, or other user-defined format files• Possibility of connecting and loading metadata from catalogue or database
	Dissemination and archival	<ul style="list-style-type: none">• libsafe is able to disseminate and audit objects without any limitation on the number of copies• Copies may be stored in different technologies and in different geographical locations



libsafe preservation platform: features

Catalogue and retrieval	Search criteria	<ul style="list-style-type: none">• Three methods for object search: Surfing the collection, simple search and advanced search• Simple search allows the user to search for text in the object name or any metadata field.• Advanced search allows the user to specify search criteria in individual metadata descriptors, and combine multiple search criteria• The search results can be filtered and sorted by any field result
	Object sheet and visualization	<ul style="list-style-type: none">• Once the object has been located, the user can access a detailed sheet of the state of preservation of it, including: name, metadata, folder and files structure, versions, stored copies and status, potential risks and actions record.• Some actions can be performed directly from the detailed object sheet: display, audit and retrieval.
	Retrieval of objects	<ul style="list-style-type: none">• The preserved material is available for single object retrieval, preservation area retrieval or entire collection retrieval.• The user always gets a copy of the object; The information preserved is kept isolated from external access, and free of risk of accidental modification.



libsafe preservation platform: features

Data management, audits, and safety	Versions, collisions and deletion	<ul style="list-style-type: none">• Metadata groups for uniqueness (e.g., bar code). In case of conflict, operator action is requested.• Metadata groups for versioning (e.g., title). In case of conflict, the object is preserved as a new version.• The descriptors in the groups can be in different metadata schemes• Preserved objects can not be deleted
	Security characteristics	<ul style="list-style-type: none">• libsafe stores information of the object, including its digital fingerprint and the location of each copy in a central database and in each of the copies.• As a result, the whole collection may be fully recovered from any of the copies in case of error.• libdata includes internal redundancy with the capability to recover data within the array even with two disk failure
	Audits	<ul style="list-style-type: none">• libsafe automatically audits the integrity of the whole collection. The user receives a report that guarantees that their objects are in perfect condition of preservation and management• Audits can be perform at disk, object and preservation area.• Additionally, the operator can perform manual audits.
	Uncommon processes	<ul style="list-style-type: none">• The data is stored so that in exceptional cases the whole collection and metadata can be retrieved directly from the preservation disks, even if the internal redundancy system of libdata is activated (unlike traditional RAID systems).

Content

1.
Masters
are
different.

They have great value. Its features high volume, heterogeneous formats and low frequency of access, make its conservation difficult.

2.
Traditional
methods are not
valid any more.

Backup tapes, CD, DVD and other offline storage systems are inexpensive but very insecure. Traditional storage and custom projects are more effective but unsustainable.

3.
Digital
Preservation is
the solution.

The solution is to apply digital preservation processes based on OAIS and ISO 14.721. However, these methodologies are multifaceted and difficult to apply in practice.

4.
libsafe is
preservation
made easy.

libsafe modeled, automates and simplifies the process of digital preservation of **complete, simple** and **sustainable** way.



Paseo de la Castellana, 153
28046 – Madrid

Tel: 91 449 08 94
Fax: 91 141 21 21
info@libnova.es

