Digital Preservation Made Easy.

LIBSAFE: DIGITAL PRESERVATION PLAN DESIGN AND MANAGEMENT

The **libsafe** platform is the solution of **libnova** for digital preservation and has been specifically designed to preserve the **master files produced during the digitization process** of historic and cultural heritage pieces, which require specific processes due to their *high value, low reading frequency and high storage volume* characteristics.

Main characteristics:

- **Global Approach**: It considers the preservation process as a whole, including all its aspects.
- **Efficient and flexible**: Driven by the most worldwide accepted preservation standard (ISO 14.721-OAIS), while at the same time keeping a practical and efficient approach derived from practical experience in real projects.
- **Look-through model**: The user has a total control over the status and location of their digital collection.
- **Active platform**: **libsafe** allows full control and easy evolution of the digital collection, so that it will always be accessible.

libsafe models and executes the following processes, whose characteristics and technical capabilities are described below:

- Metadata inclusion and management.
- Ingestion and quality control of the material to be preserved.
- Dissemination in multiple copies duly controlled and audited.
- Audit and condition control of the collection.
- Cataloguing and retrieval of preserved documents.
- Formats evolution and transformation.
- Reports (status, risks, statistics….) of the collection.

TECHNICAL FEATURES OF THE PRESERVATION PROCESSES WITH LIBSAFE

Preservation areas and plans

Unlimited preservation areas and plans, personalized independently

Preservation areas and plans include the definition of metadata formats, ingestion controls, automatic audit processes and file formats analysis and transformation to be applied to a group of digital objects.

- A preservation area includes materials digital objects and collections conceptually coherent (same kind and level of relevance, with a common structure of objects and metadata).
- A preservation plan defines the structure, metadata, ingestion controls and automatic audits for a given group of preserved objects.
- **libsafe** allows the creation of as many preservation areas as desired.
- With **libsafe** different preservation plans are possible for each preservation area; this allows the application of the optimal policy for each kind of material within the same preservation platform.

Preservable formats and objects
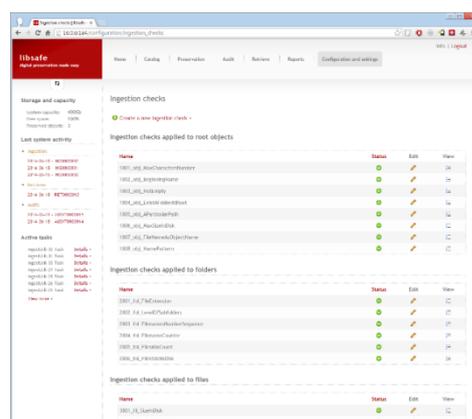
Preservation of any format of digital object.

- **libsafe** is neutral regarding the preserved content, hence it is capable to preserve any content and format among the ones usually used. Among others and not limited to: images, audio, video, texts, office documents and PDFs.
- It identifies more than 1,100 different formats. It creates a profile for each file based on DROID and verifies that real names, extensions and formats match properly.
- It checks content integrity with JHOVE.
- It provides UNICODE support for objects names, metadata and folder structure.

Ingestion and quality control

During the ingestion process, **libsafe** receives, analyzes and controls all the digital objects provided by the user, guaranteeing their quality, integrity and optimal status for the digital preservation process.

The ingestion jobs are created by the user through a very intuitive web interface that provides user friendly guidance. Based on the provided information an automatic workflow executes the following actions:

Sanitization of the ingestion material.

- **Sanitization**: temporary and operating system files deletion, permissions control, name checking for possible inconsistency conflicts in the present or future platforms.
- **Files identification and validity control**: **libsafe** identifies the file format with DROID and checks its validity with JHOVE.

Format identification and structure verification.

- **Checking the structure of digital objects**: it checks that the structure is the one defined in the preservation plan, regarding:
  - o Names of folders and files.
  - o Presence of files, folders or specific critical data.

> o   Verification of maximum and minimum size of both files and folders.
>
> ▪   **libsafe** is capable of ingesting objects generated by package generation OAIS SIP, under bagit specifications and compressed with ZIP format.

Once the process is finished the user receives an onscreen detailed report. It is also possible to receive an email with the full report.

Metadata inclusion and management

Metadata schema and representation neutrality

libsafe analyses, includes and manages standard metadata schemes as well as others designed or personalized by the user.

```xml
<?xml version="1.0" encoding="utf-8" ?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:template match="/">

  <object>
    <object_metadata>
      <xsl:for-each select="/doc/author">
        <xsl:element name="metadata">
          <xsl:element name="description"><![CDATA[author]]></xsl:element>
          <xsl:element name="value">
            <xsl:value-of select="/doc/author"/>
          </xsl:element>
        </xsl:element>
      </xsl:for-each>

      <xsl:for-each select="/doc/title">
        <xsl:element name="metadata">
          <xsl:element name="description"><![CDATA[title]]></xsl:element>
          <xsl:element name="value">
            <xsl:value-of select="/doc/title"/>
          </xsl:element>
        </xsl:element>
      </xsl:for-each>

      <xsl:for-each select="/doc/isbn">
        <xsl:element name="metadata">
          <xsl:element name="description"><![CDATA[isbn]]></xsl:element>
          <xsl:element name="value">
            <xsl:value-of select="/doc/isbn"/>
          </xsl:element>
        </xsl:element>
      </xsl:for-each>

    </object_metadata>
  </object>

  </xsl:template>
</xsl:stylesheet>
```

The platform transforms and includes standard or personalized metadata schemas.

▪   **libsafe** is implementation neutral. It stores in a database the description, semantics and value of the metadata, and can encapsulate them in any standard format for their delivery or query.

▪   Dublin Core, Marc21 and ISAD(G) schemes are available out of the box.

▪   Any other standard scheme or even user defined ones can be included.

▪   Capable of ingesting any metadata group encapsulated in XML and converting it to any compatible scheme through transformation spreadsheets XML/XSLT.

▪   It allows the creation of filters specific for particular metadata schemes that can include even Z39.50, OAI and other queries.

▪   At the time of retrieval of a copy of the digitally preserved object, it generates its associated METS, using the original metadata or the last stored and updated ones.

Collection data management

Duplicate objects detection with conflict resolution, even among different metadata schemas.

▪   **Collision Groups**: **libsafe** detects duplicated objects or conflicts in metadata contents having equivalent meaning. For instance it identifies repeated titles or repeated ISBN codes, even whit different metadata schemes.

▪   **Conflicts resolution (collisions)**: Depending on the chosen configuration the duplicated object will be rejected, a new version of it will be created or the intervention of an operator will be required to solve the conflict manually.

Dissemination and multiple copies

Multiple controlled copies, without limitation in type or location of the storage.

▪   **libsafe** makes multiple copies in the different storage systems defined, in a way which is transparent for the user. The number, target and priority of the copies are defined in the preservation plan so that that the preservation technician does not need to know the details of the implementation technology.

▪   **libsafe** is compatible with almost all the leading storage systems currently in the market, as long as they are accessible from a Microsoft Windows Server system (CIFS/SMB)..

▪   The storage systems can be different in sizes and technologies, and can be located in different geographical locations. As a matter of fact this is the recommended configuration.

▪   For each copy the integrity and digital fingerprint is stored (as MD5) as well as information of the location of the other controlled copies.

- The information is stored and organized in such a way that it is totally accessible even without using **libsafe**, in an emergency or maintenance situation.

For critical collections, **libnova** recommends keeping at least 3 copies of each preserved object, maintaining at least one in **libdata** units. For more information see the caption "*technical requirements - preservation storage*".
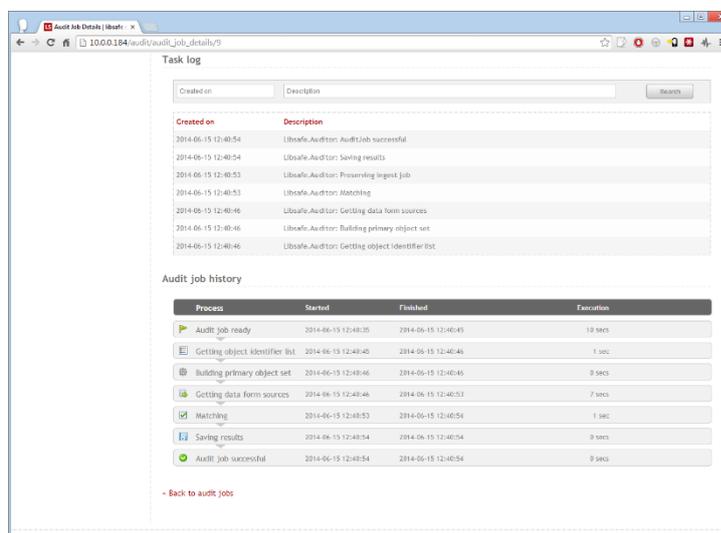
Audit and Status Control

Automatic and scheduled audit jobs are the guarantee of the integrity of the collection at a minimum effort.

During the audit jobs **libsafe** analyzes and verifies periodically and automatically that you collection(s) continues to be accessible, and keeps the same status under which it was preserved; it also checks that the collection is free of integrity and format risks.

Each copy stores integrity and digital fingerprint information of the stored objects. This information is also stored in the central database of libsafe. The audit tasks check that all the copies as well as the objects themselves are still consistent.

- **Audit types**. **libsafe** provides preconfigured automatic audits as well as automatic user defined audits and manual ones.
- **Audit scope**: An audit can be executed over a preserved object, a preservation area, a disk, or the whole collection.
- **Self-healing**: Whenever it is possible guaranteeing the solution integrity, **libsafe** corrects the detected errors; in any other case it produces a detailed report, accessible through mail or the web interface, said reports indicate the affected objects, identified risks and possible solutions, so that the use can decide the best action course.
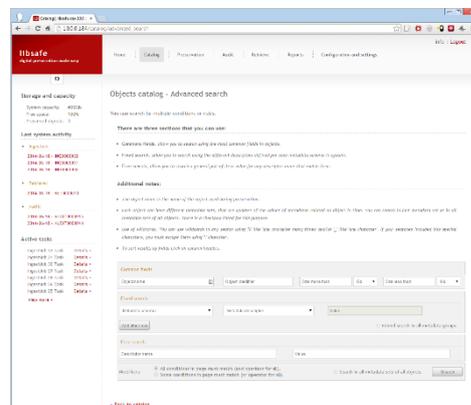
Access, catalog and retrieval

Access to the collection is possible through a catalog with search interface for pre-visualization or printing as well as retrieving specific objects.

Catalog and pre-visualization

Catalog and visualization flexible and rapid.

The catalog is the ideal solution for locating and pre-visualizing specific objects.

- **Surfing** through the collection, a specific preservation area or the whole object list.
- **Simple search** through the object name or any metadata descriptor.
- **Advanced** search through specific metadata fields, object size, name or any combination of them.
- Once a specific object is located a detailed schedule of it is presented (see "*collection reports*").

Objects retrieval

Safe retrieval: the preserved object is always out of the reach and under control.

Retrieval has been conceived to extract objects and manipulate them outside **libsafe**.

- Possibility of retrieving a single object, a preservation area or the whole collection (bulk retrieval).
- Retrieved objects are copied to a middle area accessible to the user, from which it can be moved, copied or modified.
- The original preserved object is always out of user reach, so that manipulation and or accidental erasure are avoided.
- If, once a modification is done, the user wishes to preserve the changes, a new ingestion task has to be performed. The modified object will then be preserved as a new version or as a new object. In no case the original object will be replaced

Collection report

Status reports and statistics

Flexible and detailed reports about your collection.

Reports are available for the whole collection or for a specific preservation area.

- Ingestion tasks performed report (Status and results).
- Audit tasks performed report (Status and results).
- Collection content reports: preservation areas, preserved objects, profile and format used.

Detailed sheet for each preserved object

Produced for a specific object, it includes:

- General information and object preview (name, ingestion task, preservation area, others).
- Metadata schema used and metadata content.
- Passive preservation information: Folders and files structure, format profile, disseminated copies location.

- Active preservation information: Audit and retrieval activities, transformation and evolution activities.
- Object versions (preserved objects that are stored as versions).
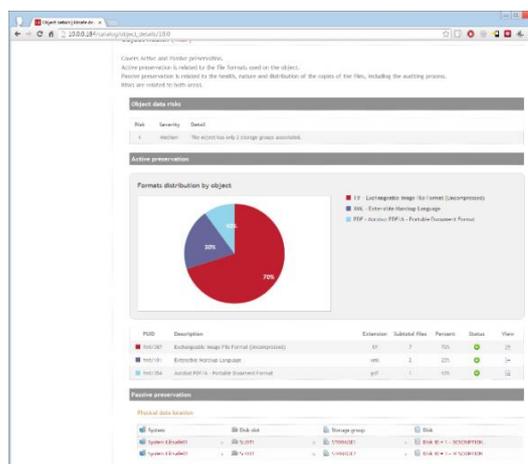
Technology Surveillance and risk reports

It shows the health status and potential risks of the preserved object. It can be executed on the whole collection, a preservation area or a specific object.

- Based on **libnova**'s knowledge database, **libsafe** produces risk reports for any given collection: number of optimal copies, obsolete or under potential obsolescence risk formats, not recognized formats.
- Through technological surveillance processes the knowledge and formats database will always be updated and available.

Note: An internet connection is needed for the Technological Surveillance functionality.

Functionality integration and update

libsafe's design allows the inclusion of new functionalities through the use of plugins. This architecture also allows the integration of **libsafe** with any other application or information system of the customer.

The user can add their own plugins. These can be developed in any programming language and must be executed under Microsoft Windows. **libnova** will provide input/ouput and API specifications to the developers.

The main points for the integration with **libsafe** workflow are:

- Call to plugins owned or designed by the customer, in the ingestion phase in order to perform advanced and customized cleaning or verification tasks.
- Call to plugins owned or designed by the customer in the metadata loading phase, in order to include customized metadata formats or schemes or to incorporate them online from a catalog or any other database .

The risk reports apply the experience and research of libnova to your collection

Extensible with customer programmable plugins.

TECHNICAL REQUIREMENTS OF THE SYSTEM

Preservation storage

libsafe stores the collection in customer owned devices, so that the control and safety are maximized. It is compatible with almost all the existing leading storage systems in the market as long as they are certified by Microsoft for use with Windows Server 2008 or later, or are certified by **libnova**.

The list of accepted devices includes, but is not limited to:

- Direct attached storage devices
- NAS servers
- SAN storage devices
- Any other storage technology accessible through CIFS/SMB

libsafe is independent from the underlying technology regarding disk types, redundancy management and security (RAID) or physical location of the devices, as long as the connectivity and bandwidth are adequate..

libsafe allows combining different types of storage under the same preservation plan; therefore it is possible to have different copies, with different technologies and in different physical locations, in an easy and transparent way, hence maximizing the security of the preserved information.

Server requierements

libsafe works on a Microsoft Windows based system. The minimum requirements are:

- Microsoft Windows Server 2008R2 with Internet Information Server (IIS).

- 64 bits quad core processor.

- RAM memory 12 GB.

- 700 GB free disk space for system and database.
  Requirements are dependent on the intensity of use, volume of data ingested and audit policies, as the processes needed to guarantee the security of the information are intensive in terms of calculations and access to both memory and disk. The above mentioned requirements are based on a standard use
  The total required disk capacity for preservation depends on the amount of information to be preserved and the number of copies.
  The intermediate storage depends on the quantity and size of ingestion or retrieval Jobs.

- Internet connection for accessing the knowledge and format databases and for downloading updates, with **libnova** remote support and maintenance agent.

libsafe can be installed in a virtualized environment and delivered as a preinstalled virtual machine. Should this configuration be chosen please contact you distributor for appropriate configuration.

Please contact a **libnova**'s official partner for more details on features and requirements for your **libsafe** digital preservation platform.