

# III Jornadas Técnicas de Archiveros Onubenses

Preservación digital de la documentación histórica



Technology changes. Information prevails.

# LLIBNOVA es líder en preservación digital.

Aportamos soluciones a las organizaciones para que **grandes volúmenes de datos valiosos estén accesibles por periodos de tiempo muy grandes.**



# Presentes en mas de 7 países



ESTADOS UNIDOS

EUROPA

LATINO AMÉRICA



¿Cuánta información  
produce la especie  
humana?

# ¿Cuánta información tenemos?



2002

5 exabytes

1 exabytes =  $2^{18}$  bytes

Fuentes:

Estudio de Marin Hillbert & Priscila López  
Estudio de la Universidad de Berkeley

# ¿Cuánta información tenemos?



2002

5 exabytes

2007

295 exabytes

1 exabytes =  $2^{18}$  bytes

Fuentes:

Estudio de Marin Hillbert & Priscila López  
Estudio de la Universidad de Berkeley



# 2002

5 exabytes



# 2011

600 exabytes



# 2020

40 zetabytes or  
5200GB by human

Fuentes:

Estudio de Marin Hillbert & Priscila López  
Estudio de la Universidad de Berkeley  
IDC

# ¿Cómo se almacena? ¿Y por qué?



Papel

8,4%



Analógico

18,34%



Digital

73,26%



¿Qué es la preservación digital?

$$ct = k(ct' + ut') = kt'(c + u) \rightarrow t' = \frac{ct}{k(c + u)}$$

$$ct' = k(ct - ut) = kt(c - u) \rightarrow t' = \frac{kt(c - u)}{c}$$

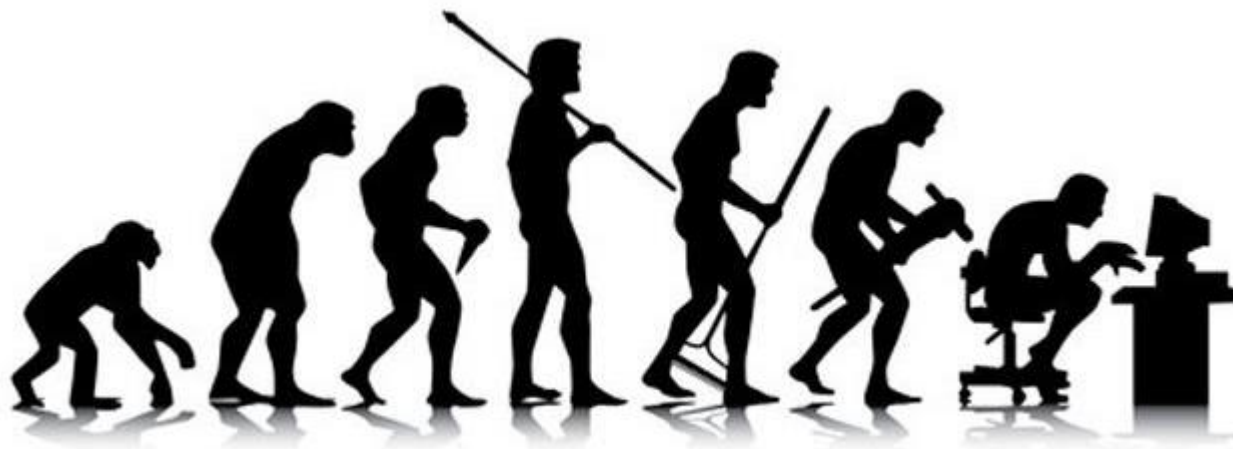
$$\frac{ct}{k(c + u)} = \frac{kt(c - u)}{c} \rightarrow k^2 = \frac{c^2}{c^2 - u^2} = \frac{1}{1 - \frac{u^2}{c^2}}$$

$$k = \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}} \Rightarrow t' = \frac{\frac{c - u}{\sqrt{1 - \frac{u^2}{c^2}}}}{c} t$$

# ¿Qué es la preservación digital?

Procesos y actividades para garantizar la reutilización futura de la información digital.

(es decir, el acceso, integridad, autenticidad y usabilidad de la misma)



# ¿Para qué preservar la info digital?

- Motivos históricos
- Motivos económicos
- Motivos legales
- ... ¿Motivos de vida o muerte?



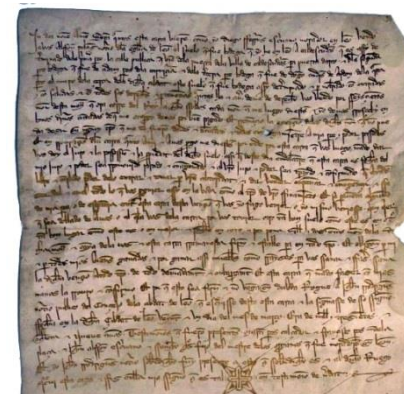
La información digital hoy es  
nuestra huella como especie y  
como cultura

**Tenemos que crear y preservar la  
memoria del futuro**

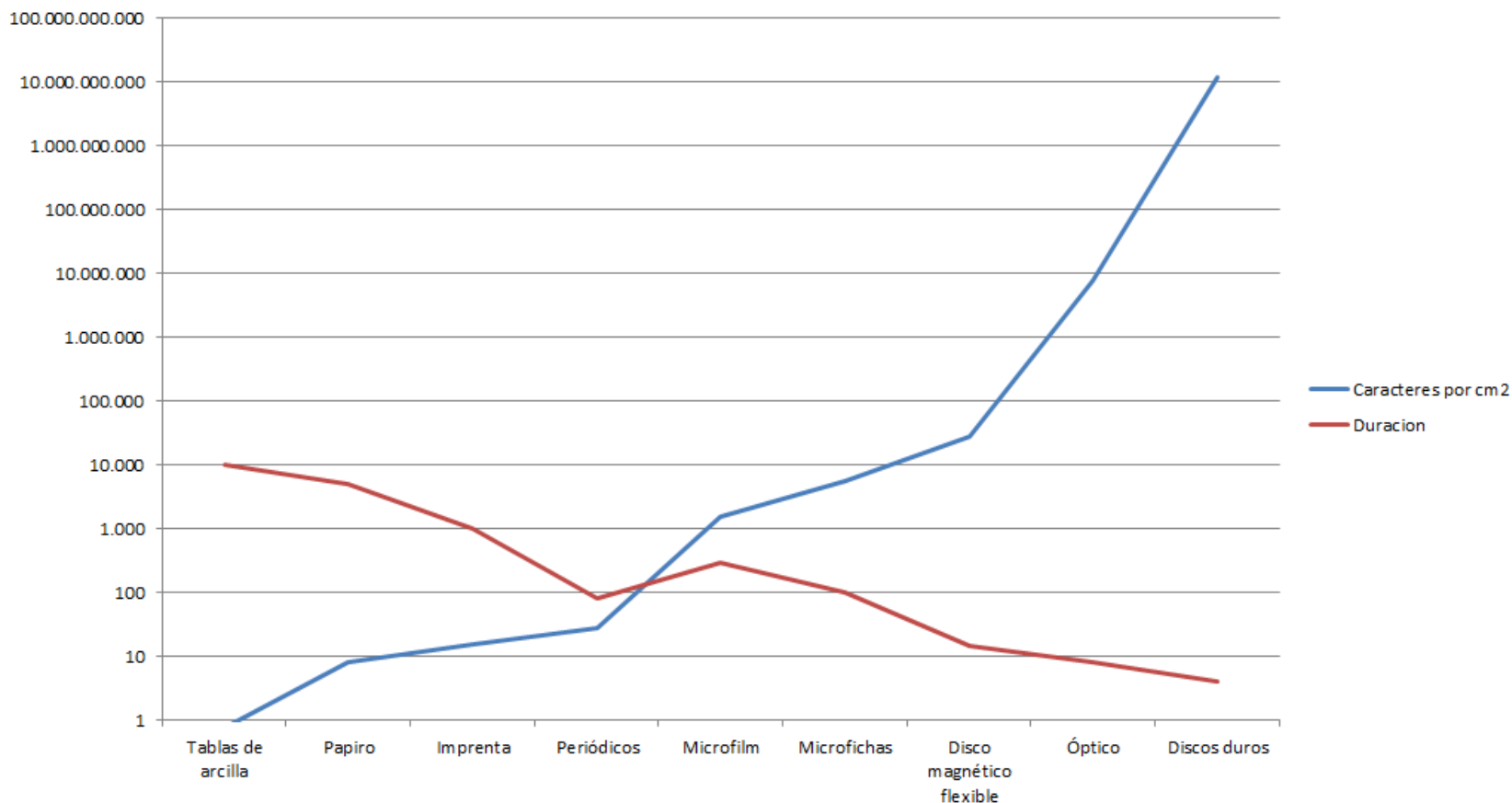


# La información digital es frágil

- El volumen es ingente
- El contenido y el soporte se separan
- Dependencias de la tecnología
- Obsolescencia tecnológica
- Mantenimiento caro
- Grandes fabricantes (Adobe, Apple, Microsoft...)



# La información digital es frágil



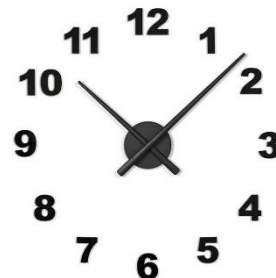
# ¿Cuánto cuesta/vale preservar?

## a) Poco

(Adaptándose a las necesidades del cliente, todos podemos preservar y hay que buscar los fondos para ello)

## b) Mucho. Incalculable.

(A priori no hay un retorno de la inversión y hay que saber hacerlo valer, no solo vas a proteger la inversión sino que te evitas el mayor problema de todos, la pérdida de “activos” en tu institución)



# ¿Y por qué es difícil?

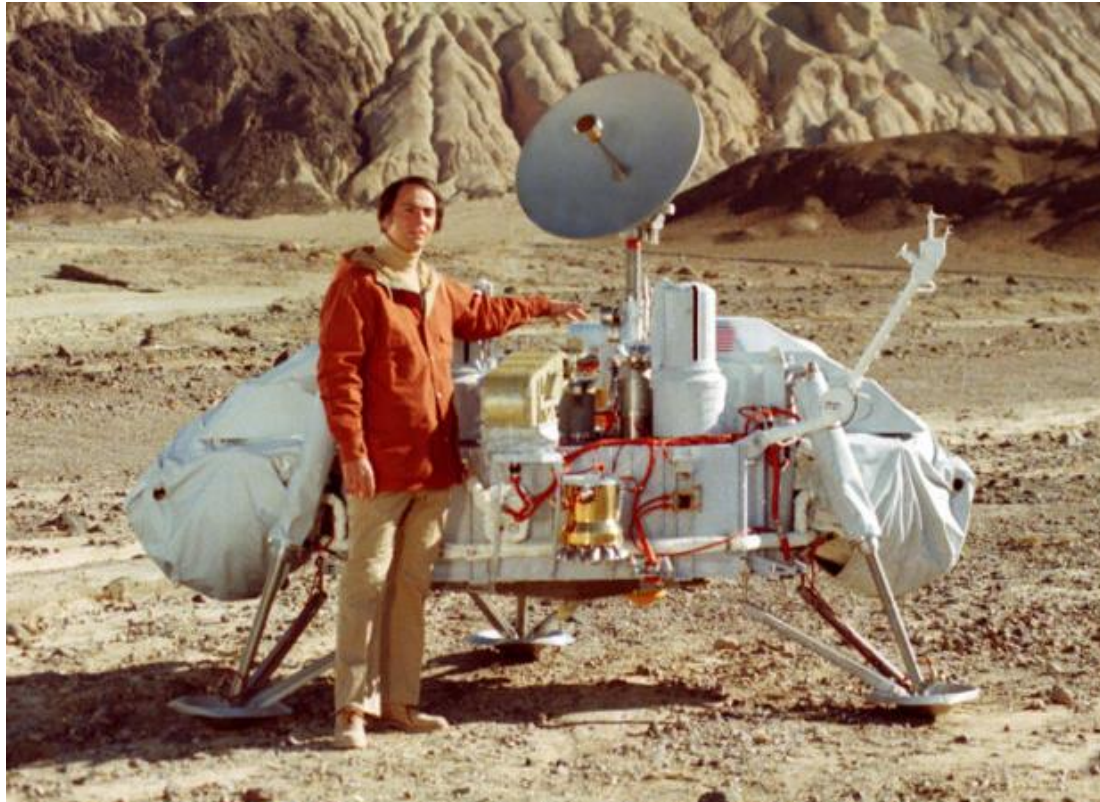


**No somos buenos midiendo riesgos  
que nos afectan en el largo plazo**



# Ejemplos de situaciones incómodas

## NASA



# La NASA y OAIS/ISO 14.721

- **1975:** La NASA envía las misiones Viking a Marte
- **1999.** El profesor Joseph Miller desea acceder a los datos. Comprueba que el 20% se ha perdido definitivamente. Archivos corruptos, 0KB, no están en sus carpetas...
- **2001:** Hay formatos que no se pueden abrir ya que los programadores habían fallecido.

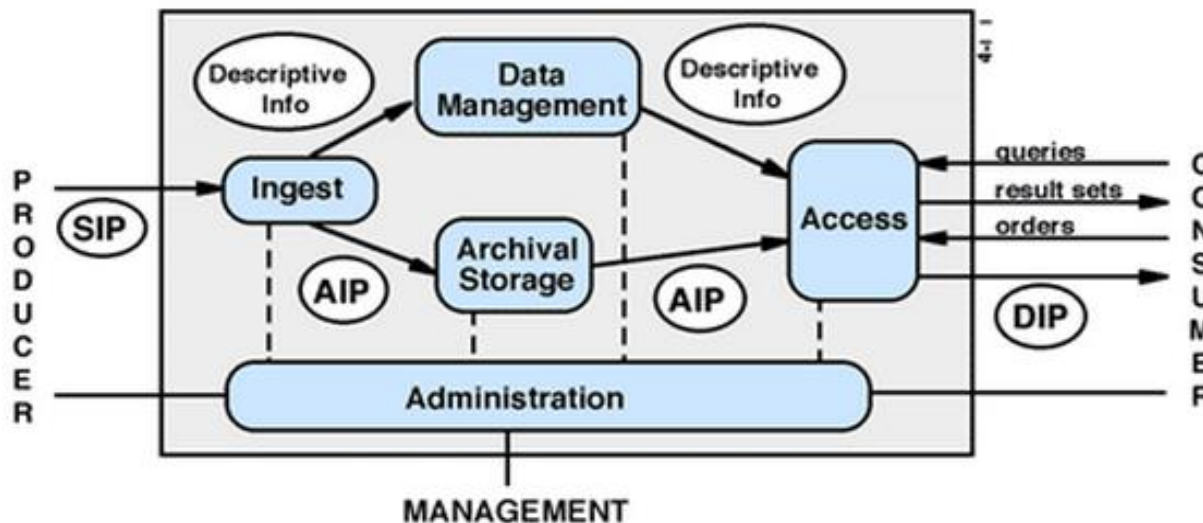


“The data were on magnetic tapes, and written in a format so old that the programmers who knew it had died”

<http://spaceflightnow.com/news/n0107/29marslife/>

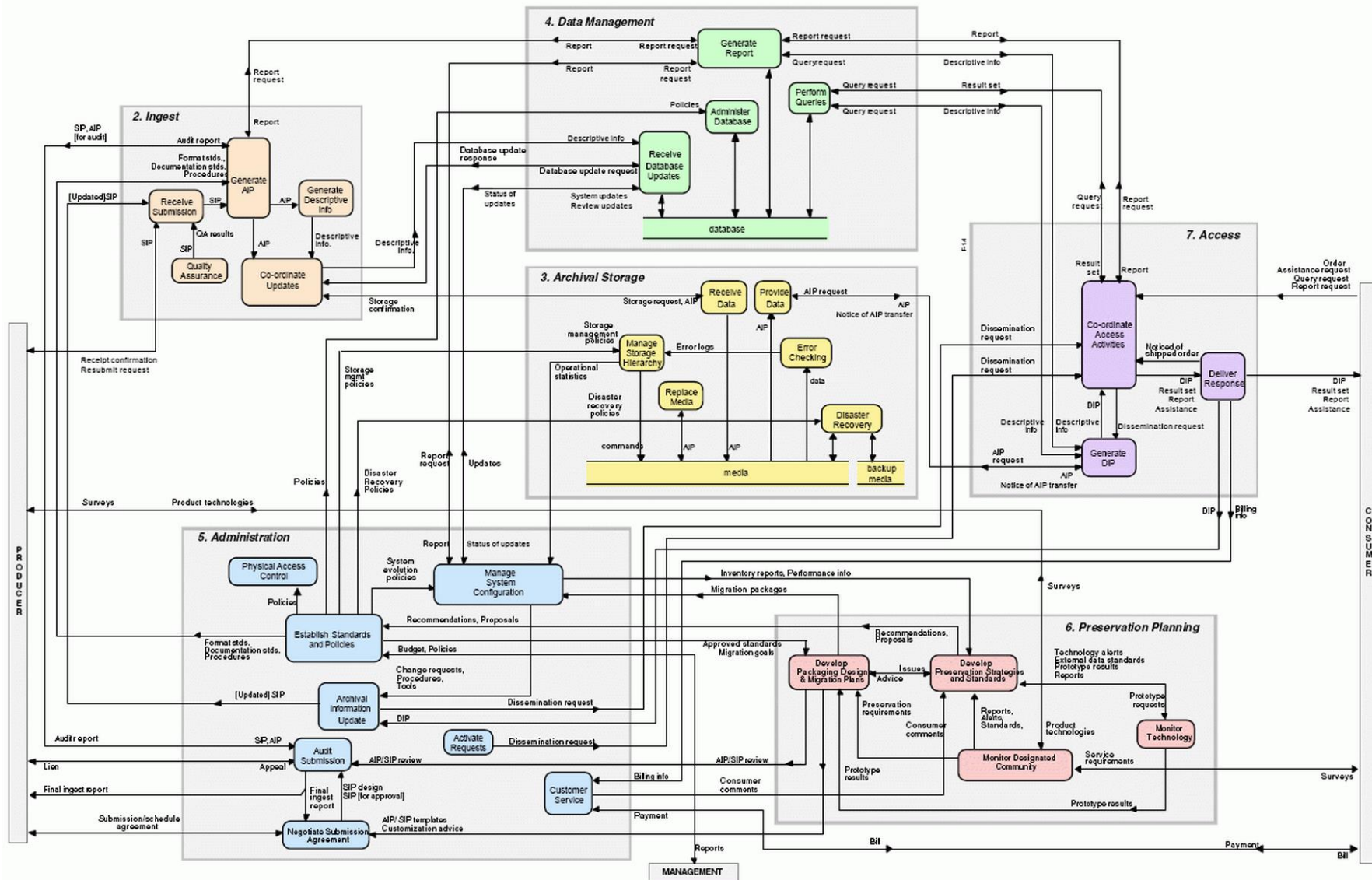
# OAIS

## The Open Archival Information System



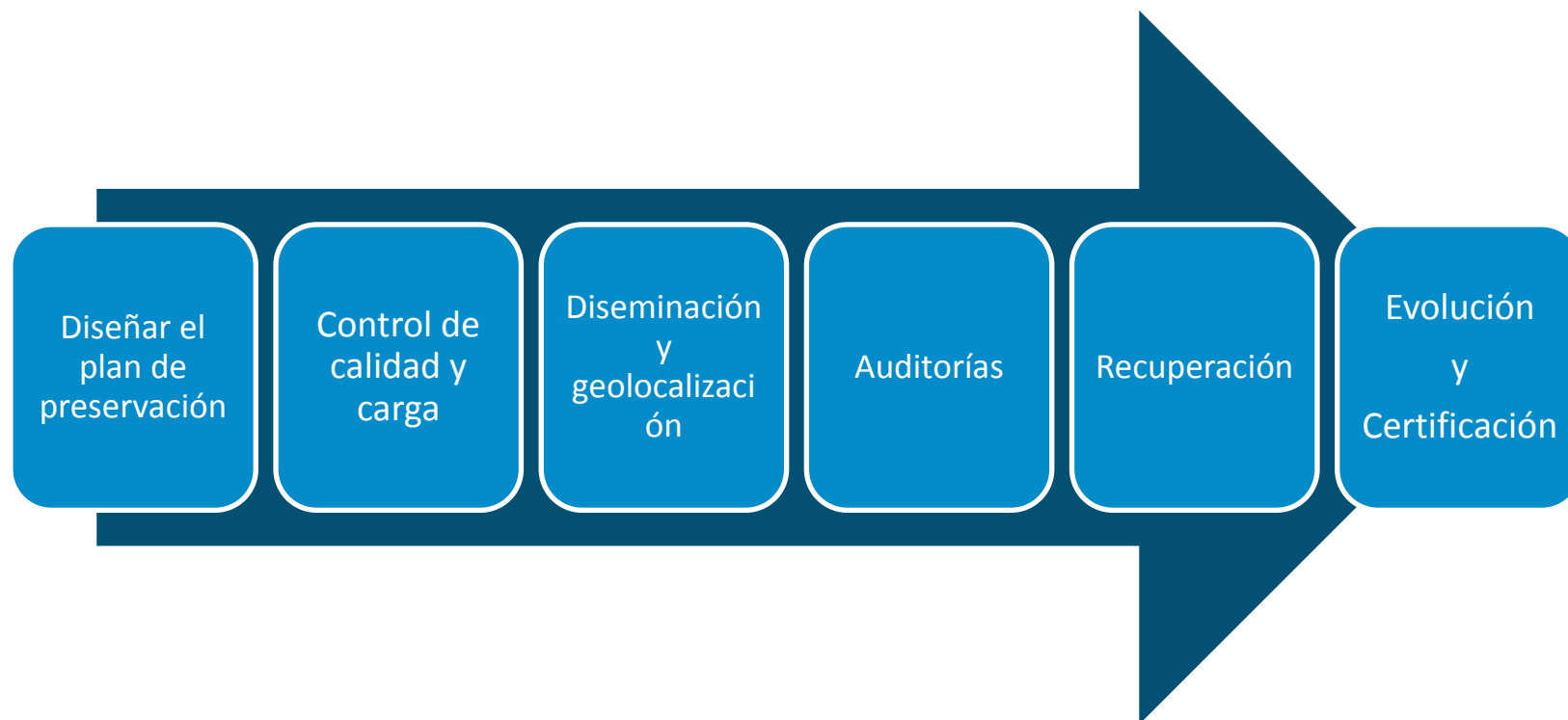
Fuente: [https://nssdc.gsfc.nasa.gov/nssdc\\_news/dec00/oais.html](https://nssdc.gsfc.nasa.gov/nssdc_news/dec00/oais.html)

Modelo conceptual destinado a la gestión, al archivo y a la preservación a largo plazo de documentos



¿Por dónde empezar?

# Puntos clave de la preservación



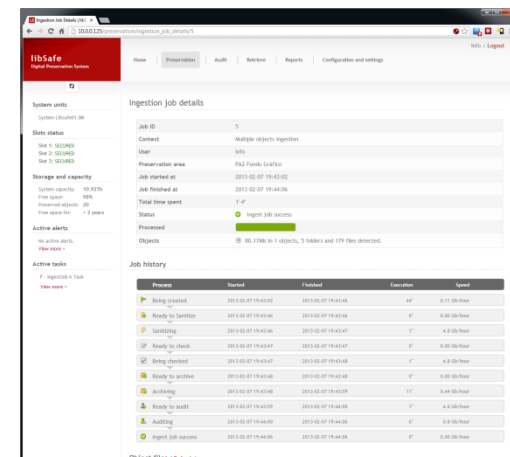
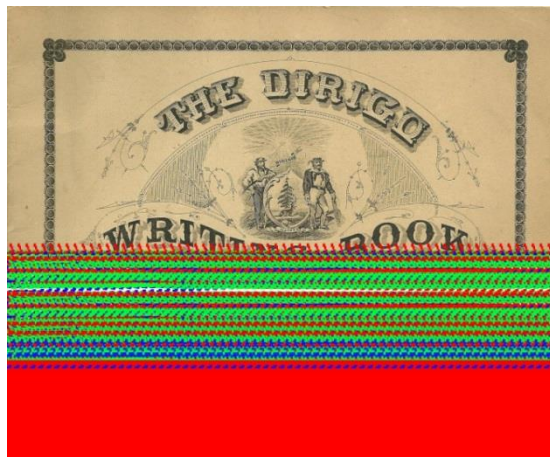
Para preservar a largo plazo hay que tomar acciones a corto plazo

# Diseñar el plan de preservación

- ¿Cómo está la colección actual?
- ¿Quién, para qué y cuando va a usar la información?
- ¿Qué preservar? TODO, lo más importante, lo que está en peor estado, la plataforma, el software, los manuales, el conocimiento de la institución...
- Definir formatos, estructuras, metadatos y soportes
- Definir derechos y seguridad de acceso a ese contenido
- Definir procesos de validación a la ingesta, auditoría y fin de vida (normalmente, para siempre)

# Control de calidad y carga

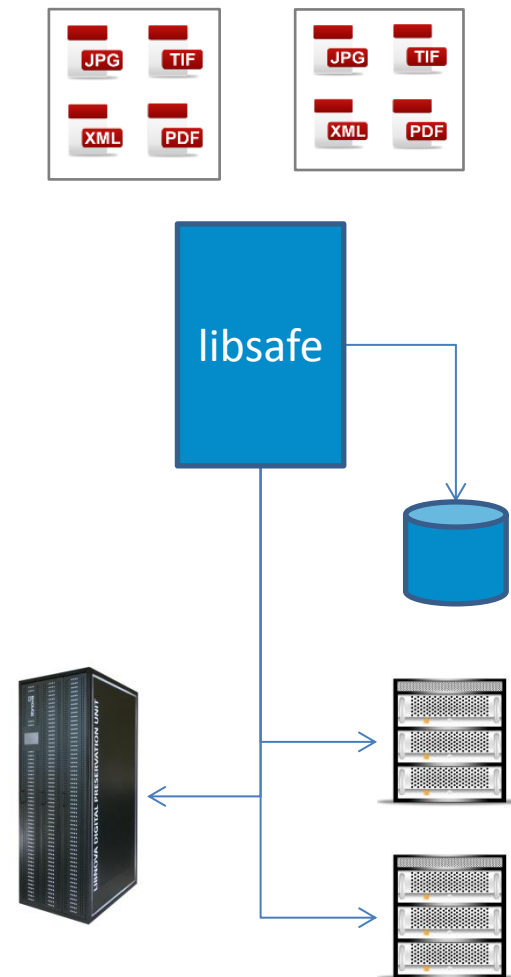
- Comprobaciones Básicas
  - Nombre, estructura de carpetas, metadatos, formatos, antivirus, huella o “HASH” y firma digital





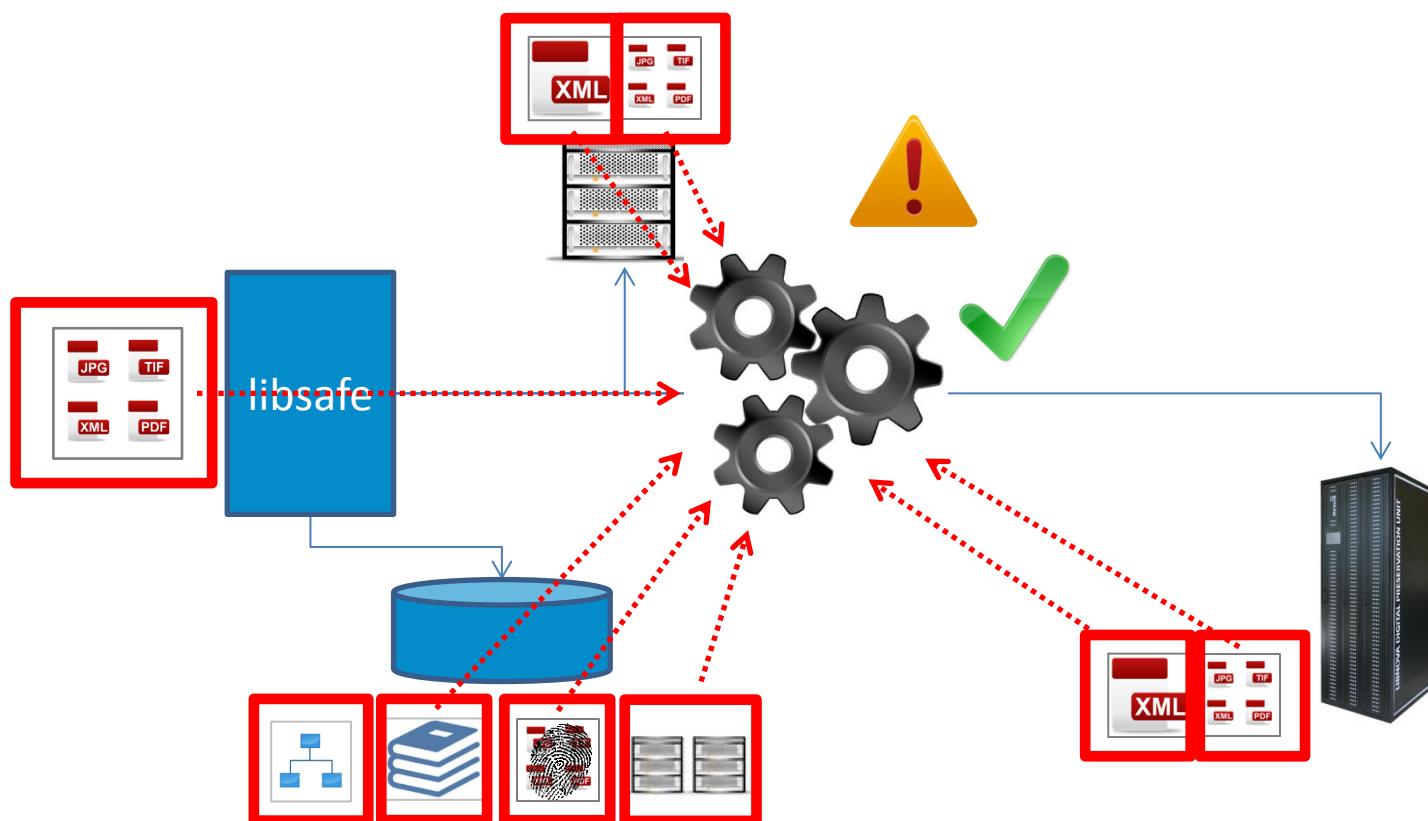
# Diseminación y copia

- Cuantas más copias (controladas) de la información, menor es el riesgo de pérdida.
- libsafe copia de forma automática a servidores, NAS o equipos libdata
- Posteriormente realiza una verificación de cada copia.



# Auditoría

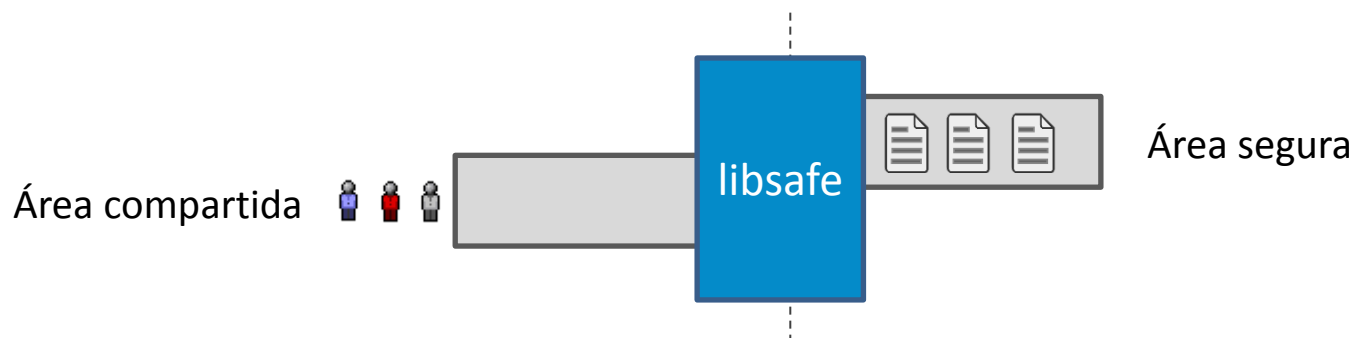
- ¿Está todo como estaba?\*



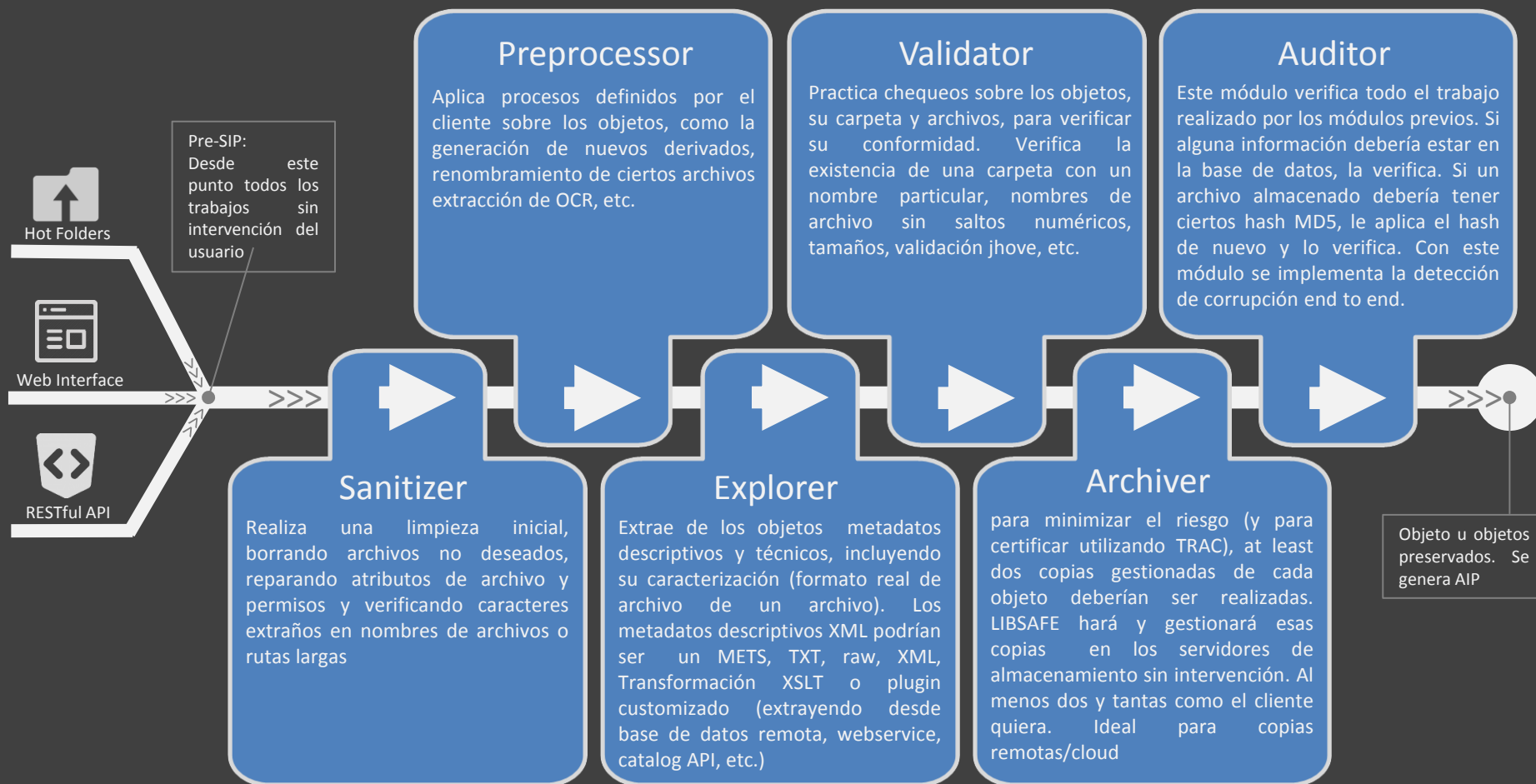
\*IT la mayoría de las veces se va a fijar en el número de ficheros, no en su integridad, ni en su formato, ni en el estado de salud de los objetos. Necesitan herramientas para ello.

# Acceso y recuperación

- Permite proteger los originales
- El usuario nunca tendrá acceso al almacenamiento real de los datos
- No pudiendo por tanto modificarlos ni dañarlo
- Los objetos se acceden por catálogo, por consulta o por URL



# Diagrama del proceso de Ingestión



# Objeto digital después de la ingestión

The screenshot shows the 'Catalog - Object details' page in the libsafe interface. The page is divided into several sections:

- Storage and capacity:** System capacity: 13GB, Free space: 79%, Preserved objects: 7.
- Current running jobs:** Ingestion: 0 jobs, Retrieve: 0 jobs, Audit: 0 jobs, Data-mover: 0 jobs.
- Last system activity:** Lists recent ingestion and audit jobs with dates and IDs.
- Active tasks:** Shows a task for 'IngestJob:5'.
- Catalog - Object details:** This is the main view of a preserved object. It includes a description: 'This is the main view of a preserved object. Here you will be able to access its files, folders, metadata details and history. On the "Object Health" section you will find information related to real storage and Active Preservation status.'
- General information:**
  - Object Name and version: ES\_GA\_36057\_AFP\_008613 (No other versions found)
  - LUOI: 1.0
  - Creation date and time: 2015-08-11 17:14:14
  - Preservation area and plan: Area: Material gráfico, using: Material gráfico, EAD Antega Sample
  - Creator: User: Administrator Method: Manual batch Ingestion.
  - Metadata schema: EAD Antega Sample
- Metadata (Hide):** A table showing metadata entries:
 

Creation date	Created by	Description
2015-08-11 17:14:14	Administrator	Initial metadata for object during preservation on IngestJob
2015-08-24 15:43:27	Administrator	Updated object metadata during preservation on IngestJob
- Files and folders structure (Hide):** A section for viewing the object's files and folders. It includes a search bar and a list of files:
 

Creation date	Created by
2015-08-11 17:14:14	Administrator
2015-08-24 15:43:27	Administrator

 Below the table, there is a search bar and a list of files:
  - jpeg
  - ES\_GA\_36057\_AFP\_008613\_00001.jpg
  - ES\_GA\_36057\_AFP\_008613\_00002.jpg
  - ES\_GA\_36057\_AFP\_008613\_00003.jpg
  - Miniaturas
  - tiff
  - ES\_GA\_36057\_AFP\_008613\_mets.xml

This screenshot shows a detailed view of the object's metadata and history. It includes sections for:

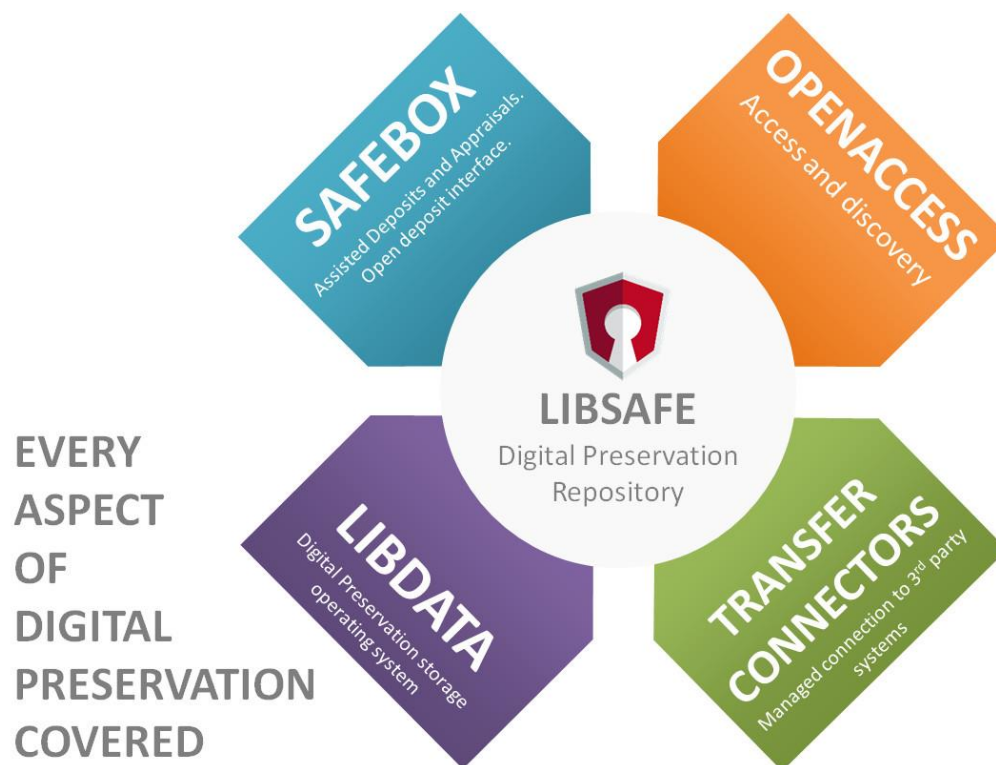
- Object Information:** Object Name, Version, LUOI, Creation Date, Preservation Area, Creator, and Metadata Schema.
- Metadata History:** A table showing the creation and updates of metadata over time.
- Files and Folders:** A list of files and folders associated with the object, including a search bar and expand/collapse buttons.
- Object Health:** A section for monitoring the object's status and preservation health.
- Events:** A log of events related to the object's lifecycle, including ingestion, audit, and retrieval.

- Creación de la información
- Metadatos (incluyendo versiones antiguas de los metadatos y evoluciones)
- Ficheros objeto (incluyendo sets antiguos de ficheros y evoluciones)
- Salud de los objetos (formatos y validaciones)
- Formatos tipo PRONOM (fmt/xxx)
- Almacenamiento físico
- Auditoría programada multipunto.
- Eventos (creación, acceso, evolución, etc.)



libdata es el almacenamiento masivo de bajo coste especialmente diseñado para preservación

Ofrece todas las prestaciones necesarias para preservación a largo plazo, sin las características que no se usan y lo encarecen



libsafe permite a las organizaciones definir y aplicar un plan de preservación

De forma sencilla, eficiente y cumpliendo los estándares

## Opciones



Cloud



On-premise



Híbrido



## **libnova Europa & LATAM**

Paseo de la Castellana, 153  
Madrid - España  
Tel: +34 91 449 08 94  
Fax: +34 91 141 21 21  
[contacto@libnova.com](mailto:contacto@libnova.com)

## **libnova USA (LIBNOVA USA)**

14 NE First Ave (2nd Floor)  
Miami, Florida 33132, USA  
Tel: +1 844-894-6532  
[jcromero@libnova.com](mailto:jcromero@libnova.com) (Country  
Manager)  
[contact@libnova.com](mailto:contact@libnova.com)

## **Contacto**

**[contacto@libnova.com](mailto:contacto@libnova.com)**

**Tel: +34 91 449 08 94**

Mi correo para preguntas

**[danielpastor@libnova.com](mailto:danielpastor@libnova.com)**

**(Daniel Pastor Ramos)**