

# SITA2020

Semana de Innovación  
Tecnológica en Archivos

Diciembre  
10-11

Preservación del documento  
electrónico: estrategias de  
identificación y control de  
formatos.

Manuela Moro Cabero  
*Universidad de Salamanca*  
moroca@usal.es



UNIVERSIDAD  
DE SALAMANCA  
CAMPUS DE EXCELENCIA INTERNACIONAL





1 -Contextualización del componente “formato” en el documento electrónico.

2 -Retos y medidas adoptadas en la identificación y selección de formatos.

3 -Herramientas con funcionalidades diversas para el control de formatos.

4 -Iniciativas *in progres*.

Preservación del  
documento  
electrónico:  
estrategias de  
identificación y  
control de formatos



1 -Contextualización del componente “formato” en el documento electrónico.

2 -Retos y medidas adoptadas en la identificación y selección de formatos

3 -Herramientas con funcionalidades diversas para el control de formatos.

4 -Iniciativas *in progres*.

Preservación del  
documento  
electrónico:  
estrategias de  
identificación y  
control de formatos.

¿Datos o documentos?

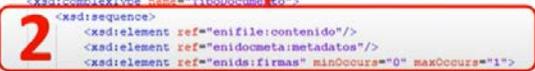
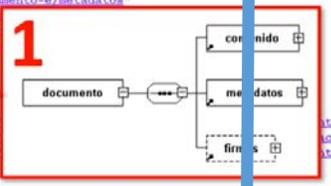
```
</xsd:choice>
</xsd:sequence>
<xsd:attribute name="Id" type="xsd:ID" use="optional"/>
</xsd:complexType>
<xsd:complexType name="TipoDocumentoIndizado">
  <xsd:sequence>
    <xsd:element name="IdentificadorDocumento" type="xsd:string"/>
    <xsd:element name="ValorHuella" type="xsd:string"/>
    <xsd:element name="FuncionResumen" type="xsd:string"/>
    <xsd:element name="FechaIncorporacionExpediente" type="xsd:dateTime"
      minOccurs="0"/>
    <xsd:element name="OrdenDocumentoExpediente" type="xsd:string" minOccurs="0"/>
  </xsd:sequence>
  <xsd:attribute name="Id" type="xsd:ID" use="optional"/>
</xsd:complexType>
<xsd:complexType name="TipoCarpetaIndizada">
  <xsd:sequence>
    <xsd:element name="IdentificadorCarpeta" type="xsd:string"/>
    <xsd:choice maxOccurs="unbounded">
      <xsd:element name="DocumentoIndizado"
        type="eniconexpind:TipoDocumentoIndizado"/>
      <xsd:element name="ExpedienteIndizado"
        type="eniconexpind:TipoIndiceContenido"/>
      <xsd:element name="CarpetaIndizada"
        type="eniconexpind:TipoCarpetaIndizada"/>
    </xsd:choice>
  </xsd:sequence>
  <xsd:attribute name="Id" type="xsd:ID" use="optional"/>
</xsd:complexType>
```



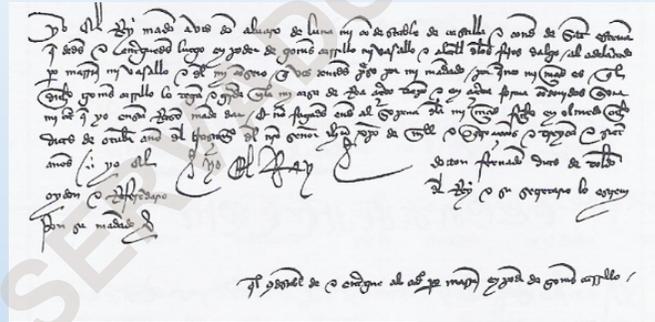
Entretenidos en dilucidar

# Datos

```
19 xmlns:xsd="http://www.w3.org/2001/XMLSchema"
20 xmlns:enids="http://administracionelectronica.gob.es/ENI/XSD/v1.0/firma"
21 xmlns:enidocmeta="http://administracionelectronica.gob.es/ENI/XSD/v1.0/documento-e/metadatos"
22 xmlns:enifile="http://administracionelectronica.gob.es/ENI/XSD/v1.0/documento-e/contenido"
23 xmlns:enidoc="http://administracionelectronica.gob.es/ENI/XSD/v1.0/documento-e"
24 xmlns:enidocmeta="http://administracionelectronica.gob.es/ENI/XSD/v1.0/documento-e/metadatos"
25 elementFormDefault="qualified" attributeFormDefault="unqualified">
26 <xsd:annotation>
27 <xsd:documentation xml:lang="es">XSD DOCUMENTO ENI (v1.0)</xsd:documentation>
28 </xsd:annotation>
29 <xsd:import namespace="http://administracionelectronica.gob.es/ENI/XSD/v1.0/documento-e/contenido" />
30 <xsd:import namespace="http://administracionelectronica.gob.es/ENI/XSD/v1.0/documento-e/metadatos" />
31 <xsd:import namespace="http://administracionelectronica.gob.es/ENI/XSD/v1.0/documento-e/firma" />
32 <xsd:element name="documento" type="enidoc:TipoDocumento">
33 <xsd:annotation>
34 <xsd:documentation xml:lang="es">El elemento "documento" podrá aparecer como elemento raíz de un documento XML o como elemento de un documento XML.</xsd:documentation>
35 </xsd:annotation>
36 <xsd:complexType name="TipoDocumento">
37 <xsd:sequence>
38 <xsd:element ref="enifile:contenido"/>
39 <xsd:element ref="enidocmeta:metadatos"/>
40 <xsd:element ref="enids:firmas" minOccurs="0" maxOccurs="1"/>
41 </xsd:sequence>
42 <xsd:annotation>
43 <xsd:documentation xml:lang="es">La firma es obligatoria para el documento administrativo electrónico v.1.0.</xsd:documentation>
44 </xsd:annotation>
45 </xsd:complexType>
46 </xsd:element>
47 <xsd:attribute name="Id" type="xsd:ID" use="optional"/>
48 </xsd:complexType>
49 </xsd:schema>
```



# Documentos



digitalizados

PDF

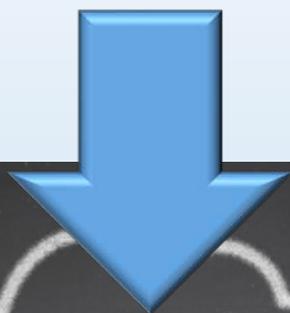


Enumerate'2017 (2019)  
52% entidades participantes (795 UE)  
No estrategia digital.  
42% disponen de estrategia escrita.  
45% no basan soluciones en estándares internac.

Estas fotos de Autor desconocido están bajo licencia [CC BY-SA-NC](https://creativecommons.org/licenses/by-sa/4.0/)



ARCHIVO  
GENERAL  
DE LA NACIÓN  
COLOMBIA



.Docx

¿Doc-dat?



# SITA2020

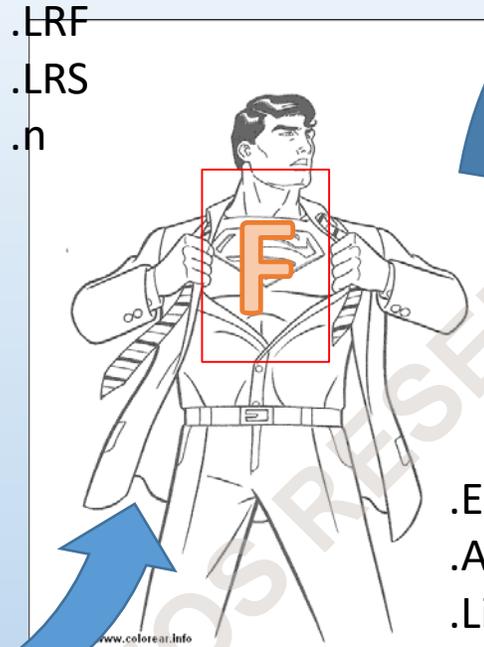
Semana de Innovación  
Tecnológica en Archivos



Diciembre  
10-11

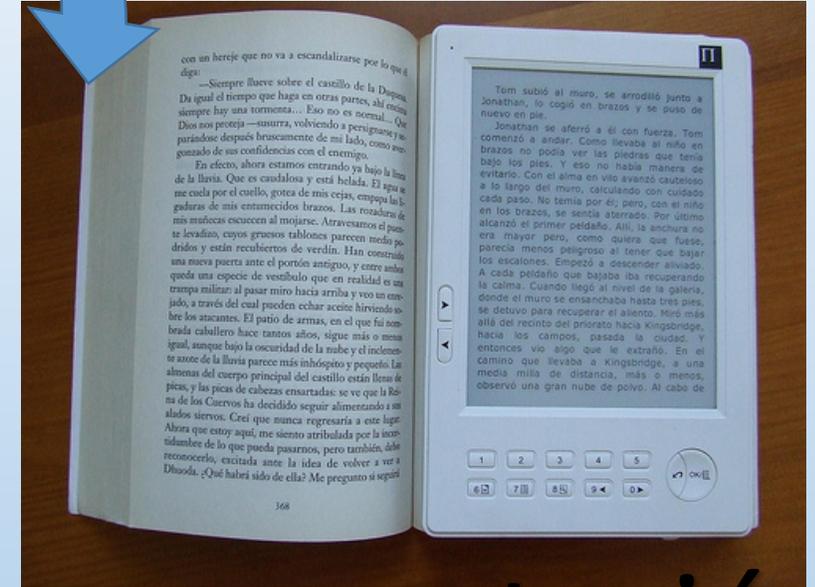
Sort by: Extension	Sort by: File type description
file extension pdf	Adobe Portable document format
file extension docx	Microsoft Word Open XML document
file extension epub	Open Publication structure e-book
file extension oxps	Microsoft OpenXPS document
file extension edoc	Electronically certified document
file extension pmd	Adobe PageMaker document
file extension pages	Pages document
file extension show	Hancom Office Hanshow presentation data
file extension _pdf	Adobe Portable Document Format
file extension eddx	Edraw project
file extension hwp	Hanword (Hangul) document
file extension btw	BarTender barcode label
file extension mso	Microsoft Organization Chart chart data
file extension pub	Microsoft Publisher document
file extension pbix	Microsoft Power BI report
file extension prn	HP PCL printer output
file extension indd	Adobe InDesign document
file extension prnx	XtraReports report
file extension lic	License file
file extension doc	Microsoft Word document (prior version 2007)
file extension mpp	Microsoft Project plan file
file extension p65	Adobe PageMaker 6.5 document

# OD



- .Epub
- .Azw
- .Lit
- .CBC
- .OPF
- .Tpz
- .n

## ¿contenido?



## representación

**FORMATO:** Conjunto de reglas (algoritmo) que define la manera correcta de intercambiar o almacenar datos en memoria.

**DATOS estructurados** cuyos principales componentes son material digital o datos, pero con un identificador unívoco, para todo él, pudiendo incluir ODs simples: ficheros de texto, de imagen, etc. u ODs compuestos, creados por combinación de un número de ODs, tales como *Web, BD estructuradas, colecciones de datos y de documentos almacenados en un ordenador.* BECKER, C. *Metamorph...* **Archivaria**, 2018, n85, p21

# realidad



Aplicaciones informáticas, bases de datos, **representaciones** de contenidos, metadatos, contenidos híbridos estructurados mediante formatos heterogéneos de imagen, datos estadísticos análisis, textuales, formularios ad hoc, firmas, etc.



Historia clínica  
electrónica

Ley 2015 de 31/01/2020

# OD y programas son ceros y unos



Datos  
Software

Decimales: etimología que es rigor merced al título, no lo poco hasta ahora lengua alguna, ni la poseer en suyo tiempo. Es efecto, para hacerse con toda propiedad etimológica en Decimales, además de contrastar la lista alfabética completa de las voces primitivas y simples, debería consignar respecto de cada una de ellas las particularidades siguientes:

1.º Su etimología inmediata, ó, mejor dicho, su origen inmediato, su etimología procedencia, más en, la indicación de la lengua de que se hallan tomados ó proceden inmediatamente, poniendo á continuación la voz de correspondencia ó la voz equivalente en dicha lengua.

2.º En qué época se halla tomado.

3.º Su significación recta ó primitiva cuando fue admitida, justificada con la cita de algún texto impreso, y un testimonio, siempre que fuere de autoridad competente.

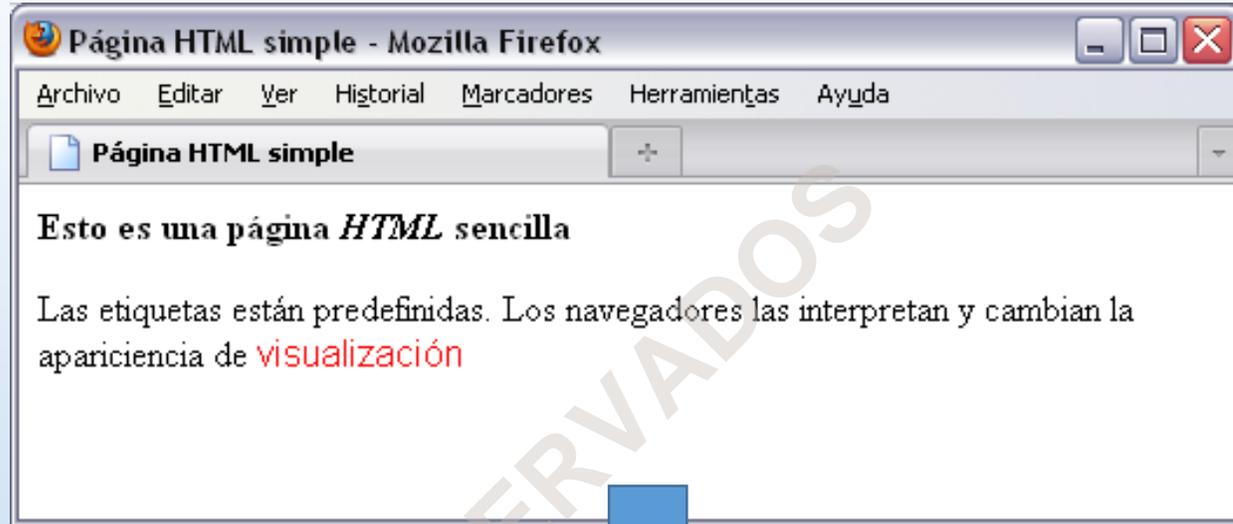
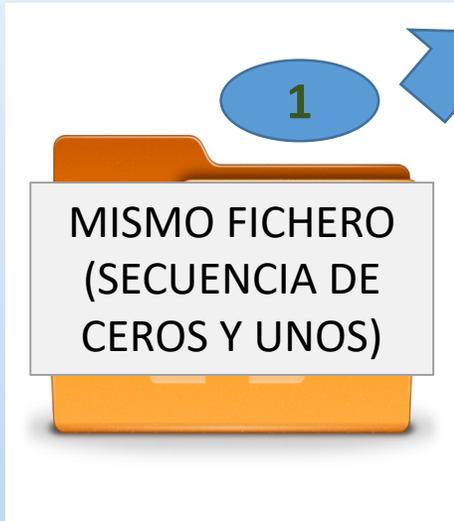
4.º La primera forma que en la pronunciación, y por escrito, tuvo aquella voz al tomarse de la lengua de origen inmediato, y las alteraciones ortográficas ó fonéticas que



- Creación
- Identificación
- Captura
- Mantenimiento
- Almacenamiento
- Recuperación
- Visualización
- Reutilización
- Etc.

DERECHOS RESERVADOS

interpretado  
por Firefox



```
1 <!DOCTYPE html>
2 <html lang="es">
3 <head>
4   <meta charset="utf-8">
5   <title>Página HTML simple</title>
6 </head>
7 <body>
8   <p>Esto es una página <i>HTML</i> sencilla</p>
9   <p>Las etiquetas están predefinidas. Los navegadores
10  las interpretan y cambian la apariencia
11  de <font face="Arial" color="#FF0000">visualización</font></p>
12 </body>
13 </html>
```

**Son los programas los que interpretan y manipulan los datos**

- Fichero en **formato CSV**

- [http://en.wikipedia.org/wiki/Comma-separated\\_values](http://en.wikipedia.org/wiki/Comma-separated_values)

The image shows two windows side-by-side. The top window is Notepad++ editing a CSV file named 'ejemplo1.csv'. The text in the file is as follows:

```
1 "kingdom_edit_account_title_text","Account Settings"␣␣
2 "kingdom_edit_account_title_text_edit_password","Change Password"␣␣
3 "kingdom_edit_account_title_text_edit_email","Change Email"␣␣
```

The bottom window is OpenOffice Calc, which has imported the CSV data into a spreadsheet. The data is organized into two columns, A and B. Column A contains the identifiers from the CSV, and Column B contains the corresponding values. Row 13 is highlighted in blue.

	A	B
1	kingdom_edit_account_title_text	Account Settings
2	kingdom_edit_account_title_text_edit_password	Change Password
3	kingdom_edit_account_title_text_edit_email	Change Email
4	kingdom_edit_account_new_email	Enter your new email.
5	kingdom_edit_account_new_password	Password needs at least four characters.
6	kingdom_edit_profile_text	Edit Profile
7	kingdom_password_prompt_enter_password	Enter your password to access these settings:
8	kingdom_lost_connection	No internet connection found.
9	kingdom_join_text	Join the Kingdom
10	kingdom_join_why_text	Choose a Kingdom name to receive additional lives, help unlocking new levels and more!
11		
12		
13		
14		

Editor de texto:  
Notepad++

2

3

Hoja de cálculo:  
Calc

# Cambio transcendental



## Formato (s)

- 1-Objetos de Datos
- 2-Programas

Características	Divergencias DE / documento convencional
Uso de símbolos	<b>Binarios codificados/decodificados</b> mediante máquina frente a comprensión alfabética del hombre.
Relación entre <b>contenido y soporte</b>	<b>Separable</b> frente a inseparable.
<b>Estructura</b>	<b>Lógica</b> frente a física
Conservación	<b>Transferencia periódica de datos</b> frente mantenimiento sólido en condiciones precisas.
Identificación	<b>Lógica</b> frente a física



1 -Contextualización del componente “formato” en el documento electrónico.

2 -Retos y medidas adoptadas en la identificación y selección de formatos

3 -Herramientas con funcionalidades diversas para el control de formatos.

4 -Iniciativas *in progres*.

Preservación del  
documento  
electrónico:  
estrategias de  
identificación y  
control de formatos.

# Observamos

.map .mso .rav .emlx .nws .pst  
.ptm .dbf .bin  
.ptt .raw .pdb .bwd .zip  
.nm3 .dcr .pdb .bwd .zip  
.mp4 .3ga .dcr .emi .tib  
.avi .flac .raf .btr .lul .rar  
.dv4 .mp3 .iiq .doc .naz .bbb  
.ughd .mpdp .hvw .pdf .nbu  
.ifv .gps .ndd .docx .xar  
.pl .wiz .exe .jar .appx  
.egg .zab .pak .egz

profusión

# Observamos

email

**.mso .rav .emlx .nws .pst**

**.mp4  
.avi  
.dv4  
.ughd  
.ifv**

*Video digital*

**.3ga  
.flac  
.mp3  
.mpdp  
.gps  
.xwm**

*Audio y  
sonido*

*cámaras*

**.raw  
.dcr  
.raf  
.iiq  
.hvw  
.nnd**

**.dbf  
.pdb  
.btr**

*Bases  
de  
datos*

*juegos*

**.bin  
.bwd  
.emi  
.lul  
.naz**

*GPS*

**.map  
.ptm  
.ptt  
.nm3**

**Variedad  
categorías**

**.pl .wiz .exe .jar  
.appx** *ejecutables*

**.doc  
.pdf  
.docx  
.pub**  
*Documento  
textual*

*backup*

**.zip  
.tib  
.rar  
.bbb  
.nbu  
.xar**

*Ficheros  
comprimidos*

**.egg .zab .pak .egz**

FileInfo		Tipos de archivo
.ITMS	Archivo de enlace de iTunes Music Store	★★★★★
.FCGI	Archivo FastCGI	★★★★★
.HABLAR CON D...	Archivo de hojas de estilo sintácticamente impresionante	★★★★★
.HDML	Archivo de lenguaje de marcado de dispositivo portátil	★★★★★
.XFDL	Archivo XFDL	★★★★★
.UCF	Archivo multimedia de WebEx	★★★★★
.PRF	Archivo de reglas PICS	★★★★★
.OTH	Plantilla HTML de OpenDocument	★★★★★
.MAPA DEL SITIO	Archivo de mapa del sitio ASP.NET	★★★★★
.DOTHTML	Plantilla de documento HTML de Microsoft Word	★★★★★
.PHP2	Página web PHP 2	★★★★★
.JSS	Archivo de hoja de estilo JavaScript	★★★★★
.WRF	Grabación de WebEx	★★★★★

# Tipologías numerosas

.ITMS	Archivo de enlace de iTunes Music Store	★★★★★
.FCGI	Archivo FastCGI	★★★★★
.HABLAR CON D...	Archivo de hojas de estilo sintácticamente impresionante	★★★★★
.HDML	Archivo de lenguaje de marcado de dispositivo portátil	★★★★★
.XFDL	Archivo XFDL	★★★★★
.UCF	Archivo multimedia de WebEx	★★★★★
.PRF	Archivo de reglas PICS	★★★★★
.OTH	Plantilla HTML de OpenDocument	★★★★★
.MAPA DEL SITIO	Archivo de mapa del sitio ASP.NET	★★★★★
.DOTHTML	Plantilla de documento HTML de Microsoft Word	★★★★★
.PHP2	Página web PHP 2	★★★★★
.JSS	Archivo de hoja de estilo JavaScript	★★★★★
.WRF	Grabación de WebEx	★★★★★

ficheros web

Formatos de sistema

DERECHOS RESERVADOS

<https://fileinfo.com/filetypes/system>



ARCHIVO  
GENERAL  
DE LA NACIÓN  
COLOMBIA



# SITA2020

Semana de Innovación  
Tecnológica en Archivos

Diciembre  
10-11

Confluencia de  
tipologías en un OD  
de Herramientas para  
su creación  
mantenimiento y  
representación



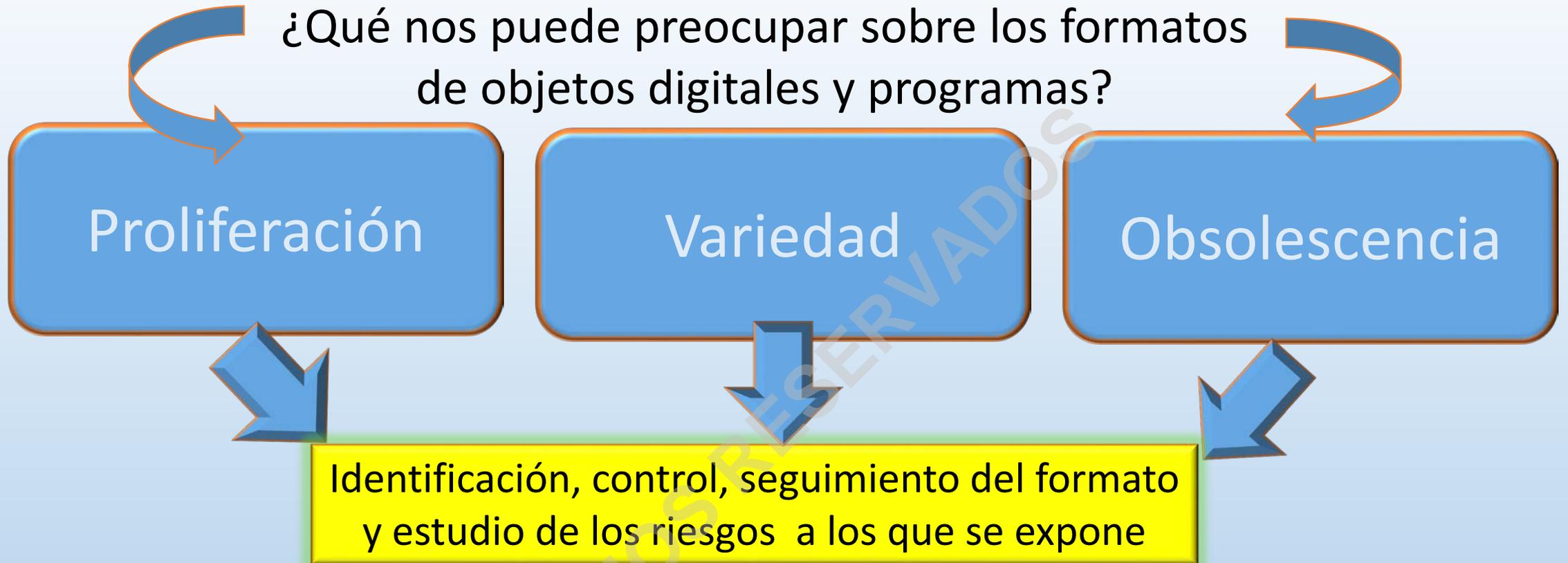
x-fmt / 44	<i>i</i> <a href="#">WordPerfect para documentos de MS-DOS / Windows</a>	6.0	doc wpd wp6 wp w60
x-fmt / 203	<i>i</i> <a href="#">Documento de WordPerfect para Windows</a>	5.2	w52 wp wpd wp5
fmt / 1424	<i>i</i> <a href="#">Documento cifrado de WordPerfect</a>	4.2	wp
x-fmt / 393	<i>i</i> <a href="#">WordPerfect para documentos de MS-DOS</a>	5,0	wp wp5 wpd w50
fmt / 949	<i>i</i> <a href="#">WordPerfect</a>	4.0 / 4.1 / 4.2	wp4 wpd
x-fmt / 394	<i>i</i> <a href="#">WordPerfect para documentos de MS-DOS / Windows</a>	5.1	wp5 wpd w51 wp
fmt / 892	<i>i</i> <a href="#">Documento compuesto de WordPerfect para Windows</a>	6 en adelante	wpd doc wp6 wp w60
fmt / 1293	<i>i</i> <a href="#">602 Documento de texto</a>		wpd wpt



Formatos  
*for ever*

# Obsolescencia y versiones

¿Qué nos puede preocupar sobre los formatos de objetos digitales y programas?



### ACCIONES IMPRESCINDIBLES

Identificar/Controlar qué debe de haber en la organización.

¿?

Controlar el riesgo vinculado al formato.

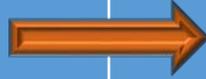
¿?

Controlar el uso y viabilidad de cada uno de los formatos.

¿?

1

Para identificar/controlar los formatos existentes en la organización



Inventario de formato  
Líneas de tiempo de la realidad

2

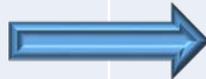
Para controlar riesgos



Criterios de selección  
Perfil de formatos  
*Aplicación a líneas de tiempo*

3

Para controlar el uso y viabilidad de cada uno de los formatos



Diversidad de herramientas con amplias funcionalidades.

Ni lo pienses



Es tarea tuya que eres informático



archiverx

RESPONSABILIDADES

Se parte de una idea

Conocer los  
formatos es un  
**requisito**  
importante



Desde la  
creación/  
captura

En la organización

Decisiones sobre:  
Creación  
Captura: acto/proceso  
**Formato**  
Forma  
Apariencia/estructura  
Firma(s)

En la  
preserva-  
ción

En la organización  
En la sociedad

Decisiones sobre:  
**1-Almacenamiento en el tiempo**  
**2-Descubrimiento**  
**3-Uso(s) y reutilización**



Fuente:

<https://pixabay.com/es/images/search/archivo/>

# Procedimiento Actividad/acto

SIA-Cclasif-NTI-tipo

- 0.1 Interoperabilidad
- 0.2 Portabilidad
- 0.3 Seguridad

## Activo informativo [objeto(s) digital(es)+ metadatos+firma(s)+índi- ces+**formato(s)**+ forma+apariencia+carpeta (s)]

Serie-Doc compuesto- NTI-  
tipo(s)\_eEMGDE

### HERRRAMIENTAS DE CONTROL

Entrada de **documentos**  
conforme a requisitos  
*Validador ENI*



*Inside Archives*

NTI-w y NTI-Política de GDE

- 1-Creación/Captura
- 2- Mantenimiento
- 3- Difusión (acceso y descubrimiento)



## 1- CONTROL DE LOS FORMATOS

Guía de aplicación de la Norma  
Técnica de Interoperabilidad

FORMATOS DE FICHEROS de contenidos	ESTÁNDARES	EJEMPLOS
Imagen /texto	Formatos de texto e imagen	.Pdf; .Png; .PdfA; .Rtf; .Tiff; .Txt; .Html, .Htm
Cartografía vectorial y sistemas de información geográfica (SIG)	Formatos de información geográfica y espacial	.gml
Compresión de ficheros	Formatos de compresión	.gz; .zip;
Contenedores multimedia	Ídem de contenedores multimedia	.avi; .mpeg; .mp4
Sonido	Ídem de sonido	.Mp3; Mpeg-1; Ogg; .Og (OGG-Vorbis)
Vídeo	Ídem de vídeo	.Mpeg-4; .Mp4; .Webm
<b>FORMATOS DE GESTIÓN DOCUMENTAL</b>		
	Estándares de gestión documental y de archivo	Normas ISO; Normas ISAD-G;.
<b>FORMATOS DE PROTOCOLOS DE COMUNICACIÓN</b>		
Correo electrónico	Tecnologías y especificaciones de transferencia de información	MIME SMTP
Páginas Web	Protocolos y estándares de intercambio de datos en entornos Webs	WS DL WS Security

Cadena de interoperabilidad	Categoría	Nombre		Tipo	Versión	Extensión	Estado
Accesibilidad multicanal integrada y segura	Autenticación Firma electrónica	Común	Formal	Abierto	1.6.3	.p7s .csig	Admitido
		CADES	ETSI TS 101 733 Electronic Signatures and Infrastructures (ESI); CMS Advanced Electronic Signatures (CADES)				



ARCHIVO  
GENERAL  
DE LA NACIÓN  
COLOMBIA



1-¿Qué hacer con otros  
ficheros y entidades que no  
están sujetas a un catálogo de  
estándares?  
2-¿Qué hacer con estos  
ficheros en 5 años?

# SITA2020

de Innovación  
en Archivos

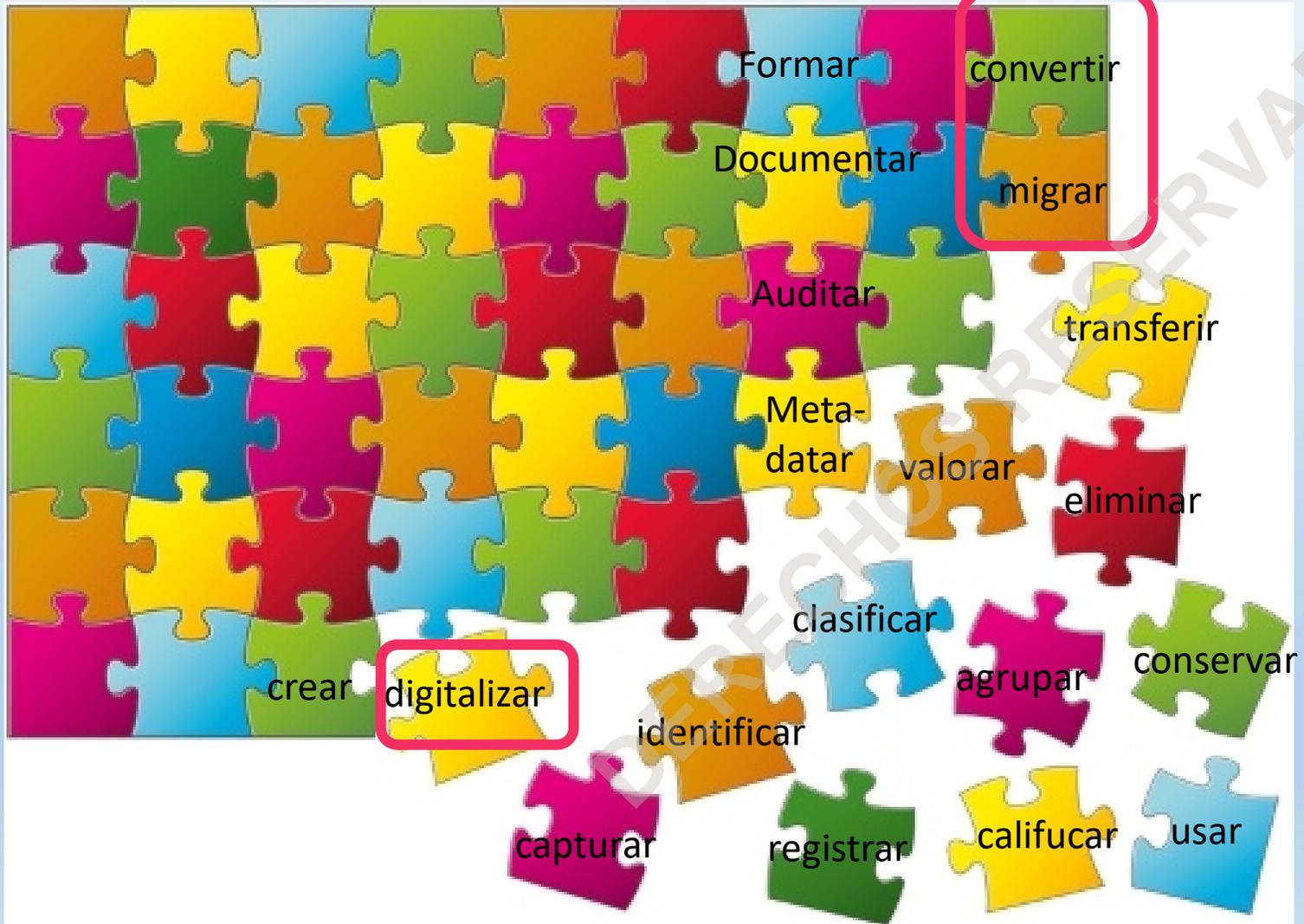


Diciembre  
10-11

DERECHOS RESERVADOS



# Archiva(liza)r



¿Las restantes piezas del puzle de “archivalizar” qué etiqueta requieren?

“Archivalizar” es construir memoria, aunque para ello... hay que ARCHIVAR Y PRESERVAR LO ARCHIVADO.



ARCHIVO  
GENERAL  
DE LA NACIÓN  
COLOMBIA

**CONSERVACIÓN FÍSICA (Archivos híbridos+ Data Center)**

+

**SEGURIDAD DE LA INFORMACIÓN**

+

**CONTROL Y TENDENCIA TECNOLÓGICA**

+

**NECESIDADES DE USO Y DE REUTILIZACIÓN DE LA INFORMACIÓN PARA COMUNIDADES ESPECÍFICAS**

+

**USABILIDAD: localización, acceso, interpretación y uso y reutilización**

+

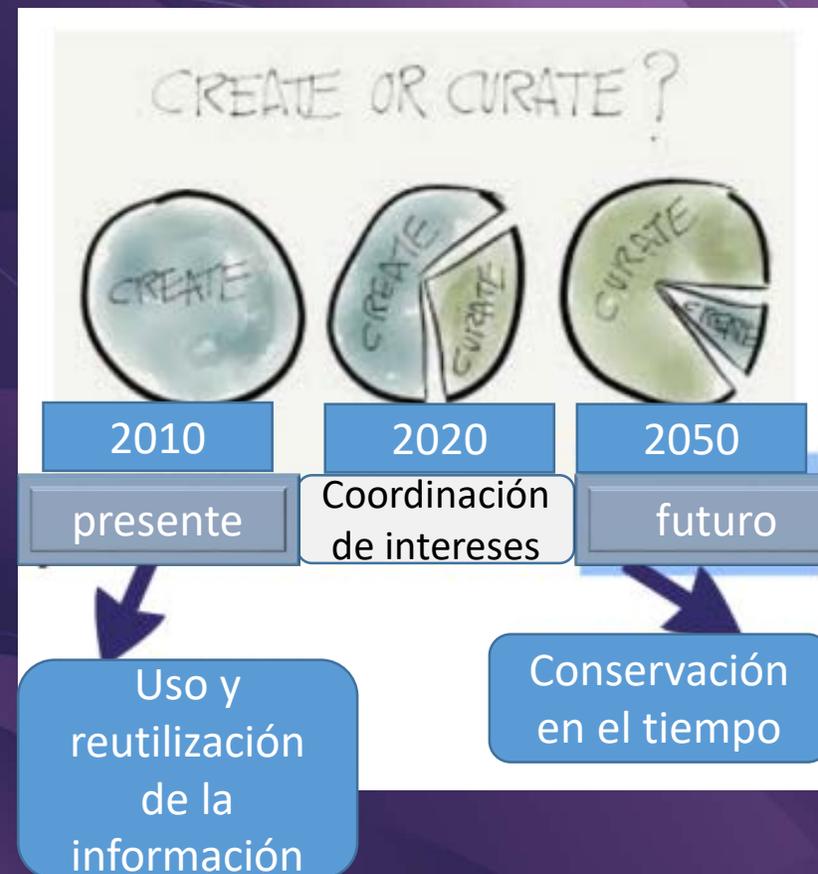
**COORDINACIÓN ESTRATÉGICA y OPERATIVA**

# Preservación digital

# SITA2020

Semana de Innovación  
Tecnológica en Archivos

Diciembre  
10-11



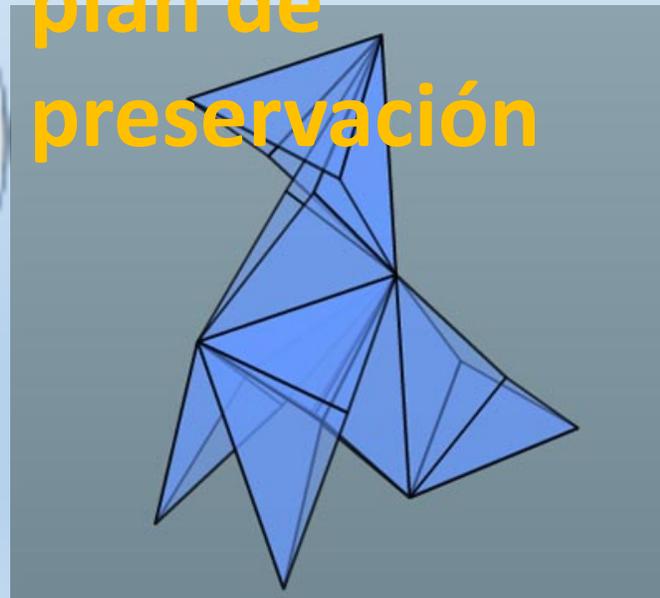


- **Plan de preservación**

- Recurso u objeto digital
- **Formato del recurso(s)**
- **Tecnología con la que ha sido creado**
- **Obsolescencia o viabilidad del formato**
- **Selección del formato**
- **Alternativa preferida, aceptada...**
- **Costes**



La **IDENTIFICACIÓN**  
y el **CONTROL** de  
los **FORMATOS**  
forma parte de un  
**plan de  
preservación**





Un **plan** exige **control** de diversos aspectos

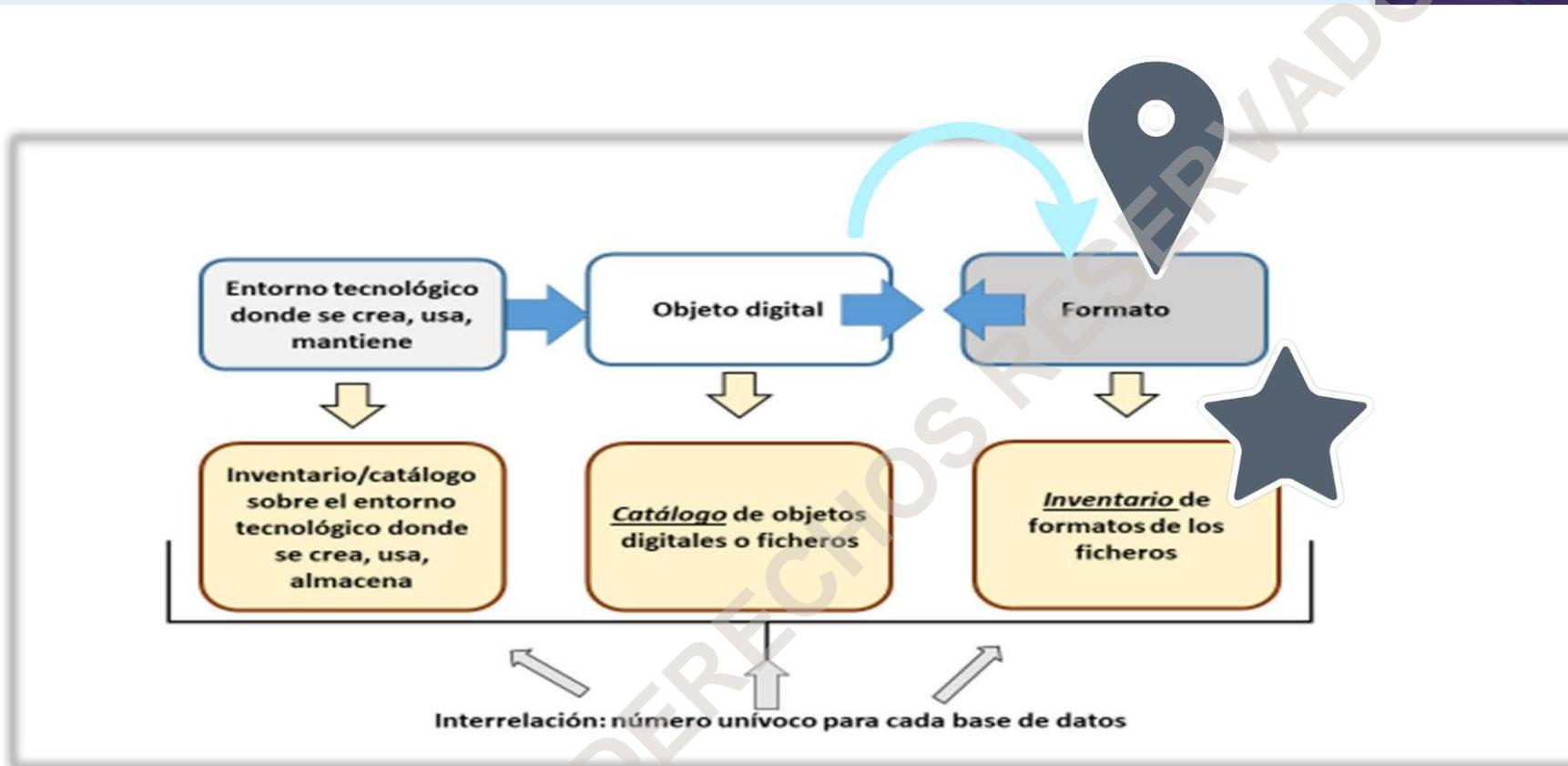


De la organización: **objetos digitales**, entorno tecnológico, **estructura de los objetos digitales**, de las necesidades de uso de los mismos por las **comunidades de usuarios**, etc.





# IMPLICA



# Un plan preservador

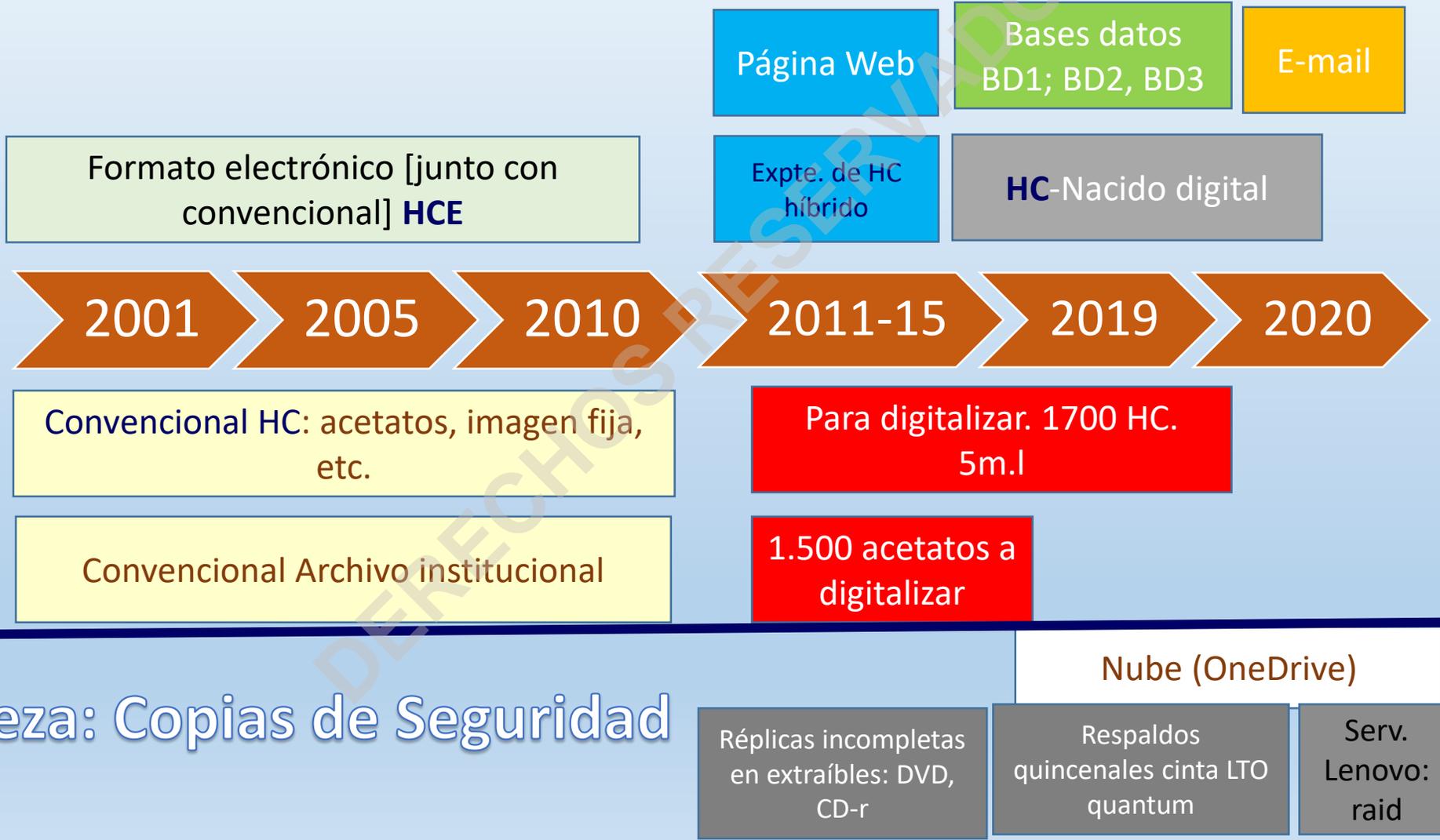
# Catálogo o inventario de OD/documentos

Nº unívoco	No de inventario formato	Acto /SIA	Código clasific.	Nombre o título	Tipo documental	Forma o naturaleza	Esencial	Dictamen -cv	Seguridad, acceso	Códigos Entorno tecnológico
										

**Otros:** datas, productor, comportamiento metadatos, necesidades usos futuros: incremento, modificación, consulta, visualización; volumen, NTI...

# Línea de tiempo de documentos/tecnología

Soportes extraíbles requieren atención: Dvd-r, CD-r; DVD r-w; CD r-w; capas simple o doble. LTO; discos magnéticos flexibles 35", memorias flash, discos ópticos, disquetes 3,1/2



Fijeza: Copias de Seguridad



2

Acciones del gestor de recursos digitales	Instrumentos y aplicaciones
<ul style="list-style-type: none"><li>Identificación de los recursos digitales existentes.</li></ul>	Catálogo de recursos digitales, líneas de tiempo
<ul style="list-style-type: none"><li><b>Identificación/control de formatos</b> existentes en la organización.</li></ul>	 <b>Inventario de formato</b> <b>Mapas gráficos o líneas de tiempo</b>



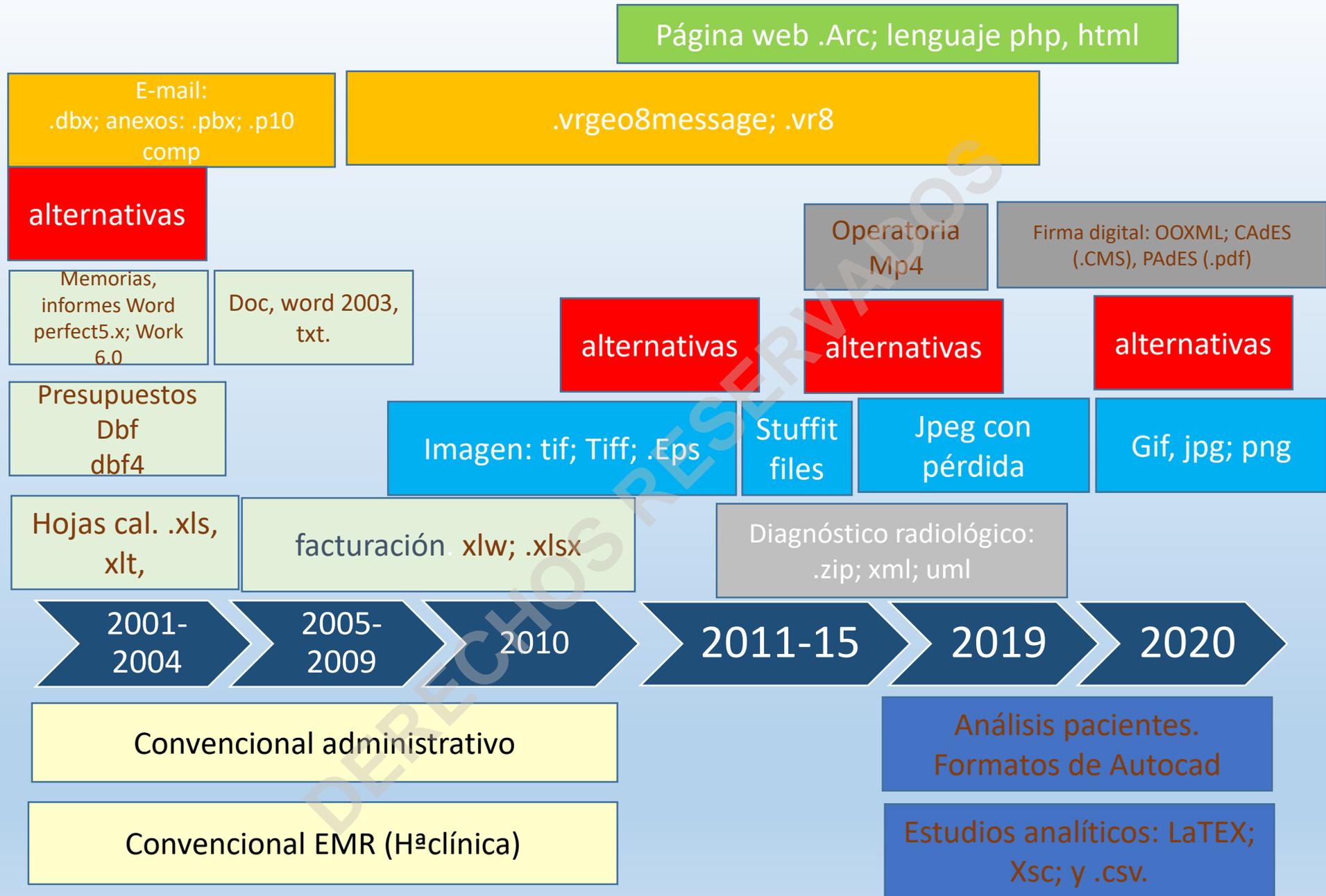
## Inventario de formatos

Nº unívoco	Formato	versión	Datas-CV	SIA/OD	Abierto/ Cerrado/ Semi-	ENTORNO TECNOLÓGICO/	Opciones de preservación: viabilidad
	Nombre extenso y extensión						<ul style="list-style-type: none"><li>• Convertible</li><li>• Se abre con</li><li>• Se consulta o visualiza con, se exporta, se importa</li><li>• Se Archiva</li></ul>

mínimos

DERECHOS RESERVADOS

### Conocimiento riguroso de los formatos existentes



Mapa gráfico de formatos (línea de tiempo)



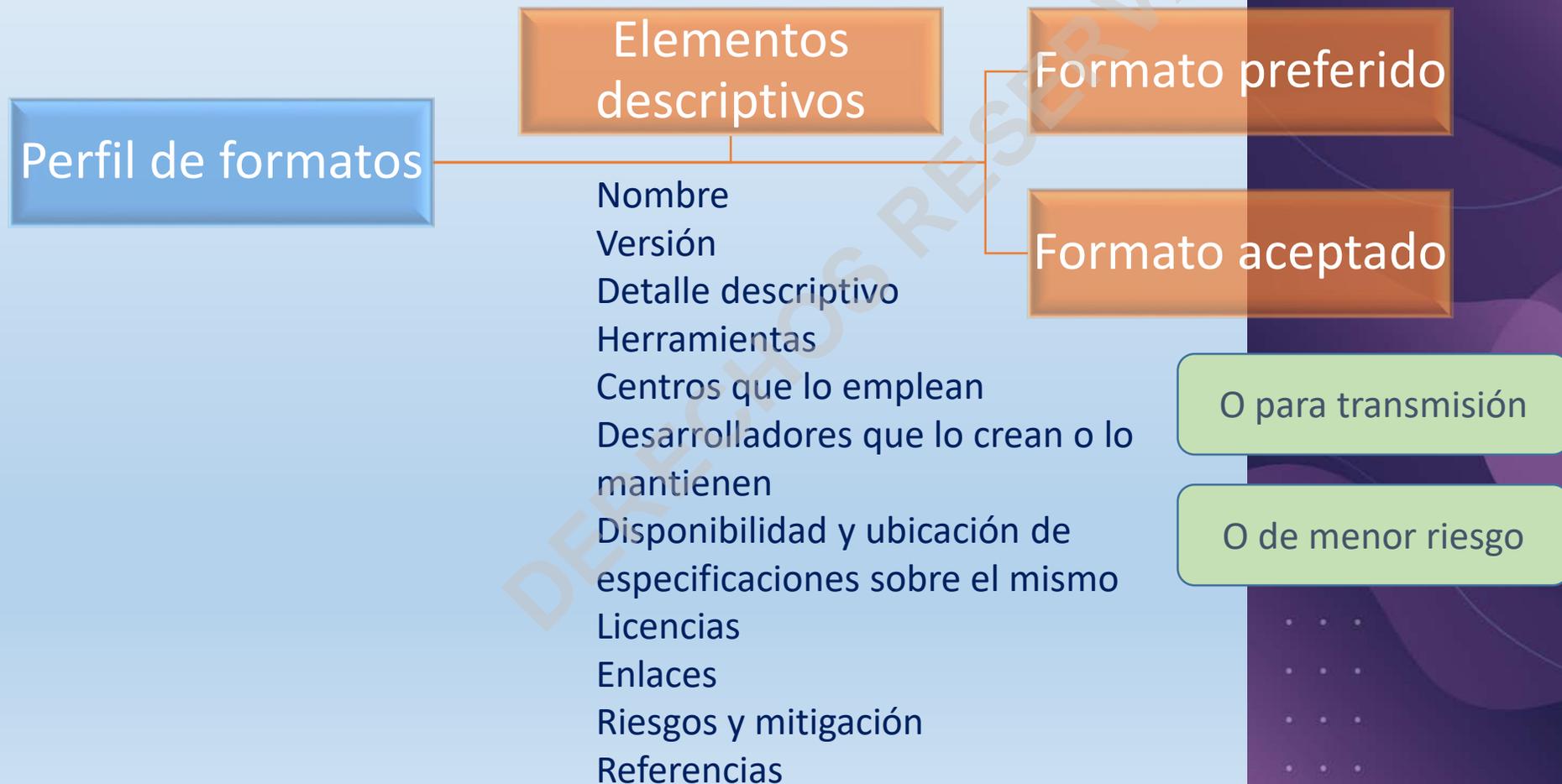
Acciones del gestor de recursos digitales	Instrumentos y aplicaciones
<ul style="list-style-type: none"><li>Identificación de los recursos digitales existentes</li></ul>	Catálogo de recursos digitales Línea de tiempo OD
<ul style="list-style-type: none"><li>Identificación de formatos existentes en la organización.</li></ul>	Inventario de formato Mapa gráfico de formatosw
<ul style="list-style-type: none"><li>Conocimiento de propiedades del formato.</li><li>Valoración y selección de formatos.</li><li>Control del riesgo vinculado a viabilidad de formatos.</li></ul>	<b>Matriz del formato</b>  Factores; Criterios e indicadores ante/para la selección  <b>Perfil de formatos</b>

1-Nombre completo (tomado del estándar si es el caso) y otros nombres con los que se conoce	4-Principales bibliotecas y repositorios o dominios que lo emplean	7-Compañía o persona física desarrollador del formato, originalmente	10-Breve información sobre problemas de patentes / licencias	13-Mitigación de los principales riesgos
2-Breve descripción (contenido clave a conocer de él, por ejemplo, versiones)	5-Tipos Mime aplicables (información sobre el contenido de un Archivo: audio, aplicación, mensaje, modelo, texto, multiparte)	8-Compañía o persona física desarrollador del formato que lo mantiene en la actualidad	11-Enlaces relacionados con datos claves del mismo (sitios web que lo describen, documentación, etc.)	<b>14-Referencias</b>  Enlace a algunos de los Mime Types más importantes: <a href="https://developer.mozilla.org/es/docs/Web/HTTP/BasicsofHTTP/MIMEtypes/Lista_completa_de_tipos_MIME">https://developer.mozilla.org/es/docs/Web/HTTP/BasicsofHTTP/MIMEtypes/Lista_completa_de_tipos_MIME</a>
3-Herramientas	6-Extensiones de archivo aplicables	9-Disponibilidad y ubicación de especificaciones (direcciones URL directas si están disponibles)	12-Riesgos vinculados, factores, etc. si se conocen	Enlace a listado General Media Types: <a href="http://www.iana.org/assignments/media-types/media-types.xhtml">http://www.iana.org/assignments/media-types/media-types.xhtml</a>  Tipos comunes de “medios” en Internet: <a href="http://www.htmlquick.com/es/reference/mime-types.html">http://www.htmlquick.com/es/reference/mime-types.html</a>

**COMPONENTES DE UNA MATRIZ DE FORMATOS (Harvard University)**



# Perfil de formatos



NARA: <http://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html#scannedtext>

1

**Simbolo clave**

Formatos preferidos

aceptables

Aceptables para formatos de transferencia inminentes

## Diseño asistido por computadora (CAD)

Diseño asistido por computadora (CAD): los formatos CAD son archivos de gráficos vectoriales que se basan en una computadora multidimensional destinada a su uso en diseño de ingeniería y fabricación. Los programas crean proyecciones de superficie de objetos bidimensionales y tridimensionales.

**Formatos preferidos**

Formatos preferidos	Especificaciones de formato
3D extensible (X3D)	ISO / IEC 19775-1: 2008: ( <a href="http://www.web3d.org/files/specifications/19775-1/V3.2/index.html">http://www.web3d.org/files/specifications/19775-1/V3.2/index.html</a> )
	ISO 10303-21: 2002: ( <a href="http://www.iso.org/iso/home/standards/catalogue_tc/">http://www.iso.org/iso/home/standards/catalogue_tc/</a> )

LIBRARY OF CONGRESS:

<https://www.loc.gov/preservation/resources/rfs/TOC.html>

ii. Obras Textuales - Digital		
	Privilegiado	Aceptable
<b>A. Características técnicas</b>	Codificación de caracteres, en orden descendente de preferencia: <ol style="list-style-type: none"> <li>1. UTF-8, UTF-16 (con BOM), US-ASCII</li> <li>2. ISO 8859</li> </ol>	Otras codificaciones de caracteres no incluidas en la sección Preferidas
<b>B. Formatos, en orden de preferencia</b>	<ol style="list-style-type: none"> <li>1. Formatos de marcado basados en <u>XML</u>, con DTD / esquema incluido o accesible, hoja (s) de estilo de presentación XSD / XSL y codificación de caracteres explícitamente establecida               <ol style="list-style-type: none"> <li>a. <u>Cumple con EPUB3</u>. (Otras versiones de EPUB también son formatos preferidos, pero EPUB3 es el más común).</li> <li>b. <u>BITS</u> (paquete de etiquetas de intercambio de libros) versión 2.0</li> <li>c. Otros DTD / esquemas de libros ampliamente utilizados (por ejemplo, TEI, DocBook, etc.)</li> </ol> </li> <li>2. Formatos de diseño de página               <ol style="list-style-type: none"> <li>a. <u>PDF / UA</u> (compatible con ISO 14289-1)</li> <li>b. <u>PDF / A</u> (compatible con ISO 19005)</li> <li>c. <u>PDF</u> (la más alta calidad disponible, con características como texto con capacidad de búsqueda, fuentes incrustadas, compresión sin pérdida, especificación de espacio de color independiente del dispositivo, etiquetado de contenido; incluye formatos de documento como <u>PDF / X</u>)</li> </ol> </li> </ol>	<ol style="list-style-type: none"> <li>1. Otros formatos estructurados o de marcado               <ol style="list-style-type: none"> <li>a. <u>XHTML</u> o <u>HTML</u>, con declaración DOCTYPE y hoja (s) de estilo de presentación</li> <li>b. Formatos de documentos basados en <u>XML</u> (ampliamente utilizados y documentados públicamente), con hojas de estilo de presentación, si corresponde. Incluye <u>DOCX / OOXML 2012</u> (ISO 29500), <u>ODF</u> (ISO / IEC 26300) y <u>OOXML</u> (ISO / IEC 29500).</li> <li>c. <u>SGML</u>, con DTD incluido o accesible</li> <li>d. Otros formatos no propietarios basados en XML, con hojas de estilo de presentación</li> <li>e. Formatos basados en XML que utilizan DTD o esquemas patentados, con hojas de estilo de presentación</li> </ol> </li> <li>2. Formatos de diseño de página</li> </ol>



ARCHIVO  
GENERAL  
DE LA NACIÓN  
COLOMBIA

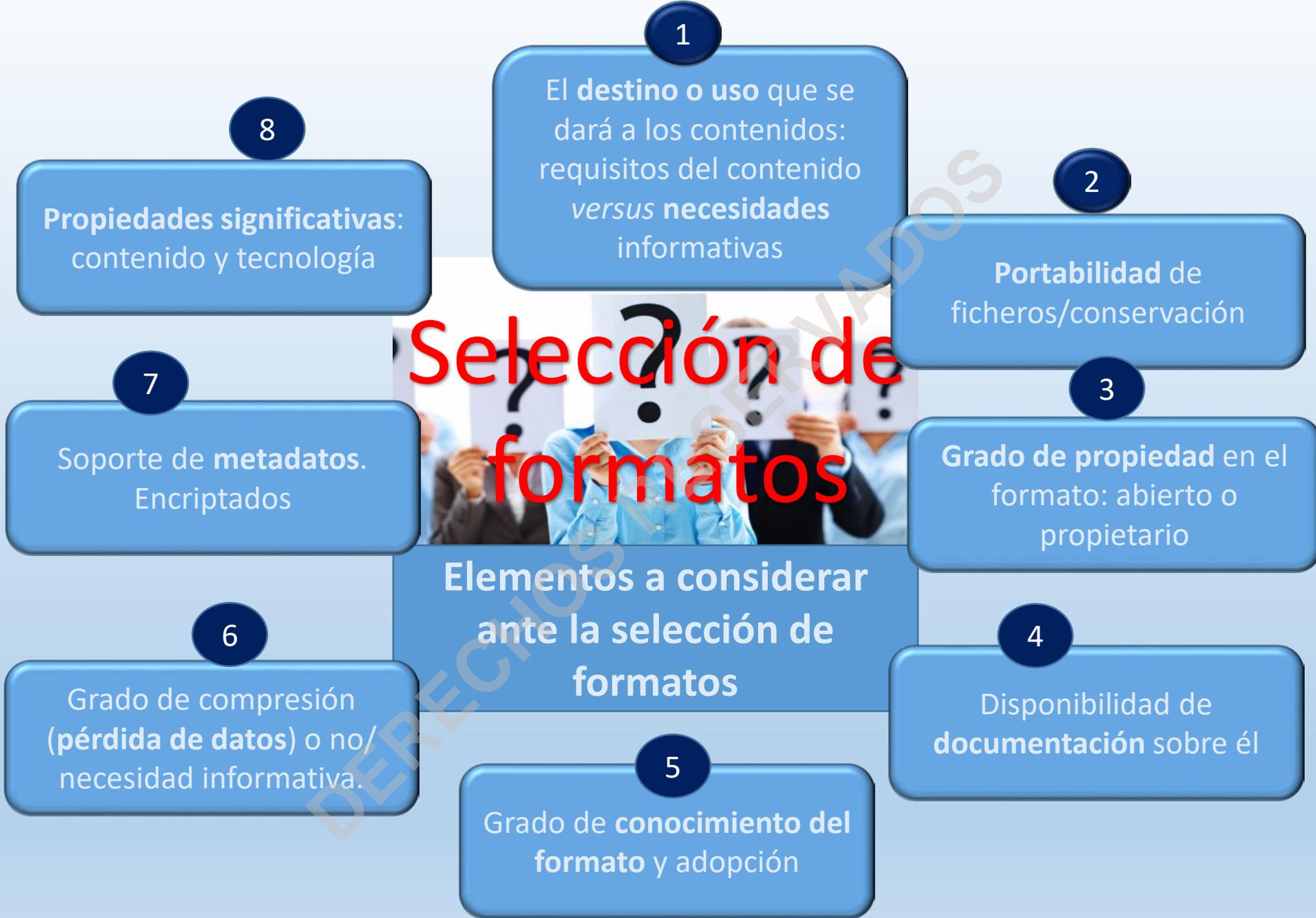
# SITA2020

Semana de Innovación  
Tecnológica en Archivos

Diciembre  
10-11

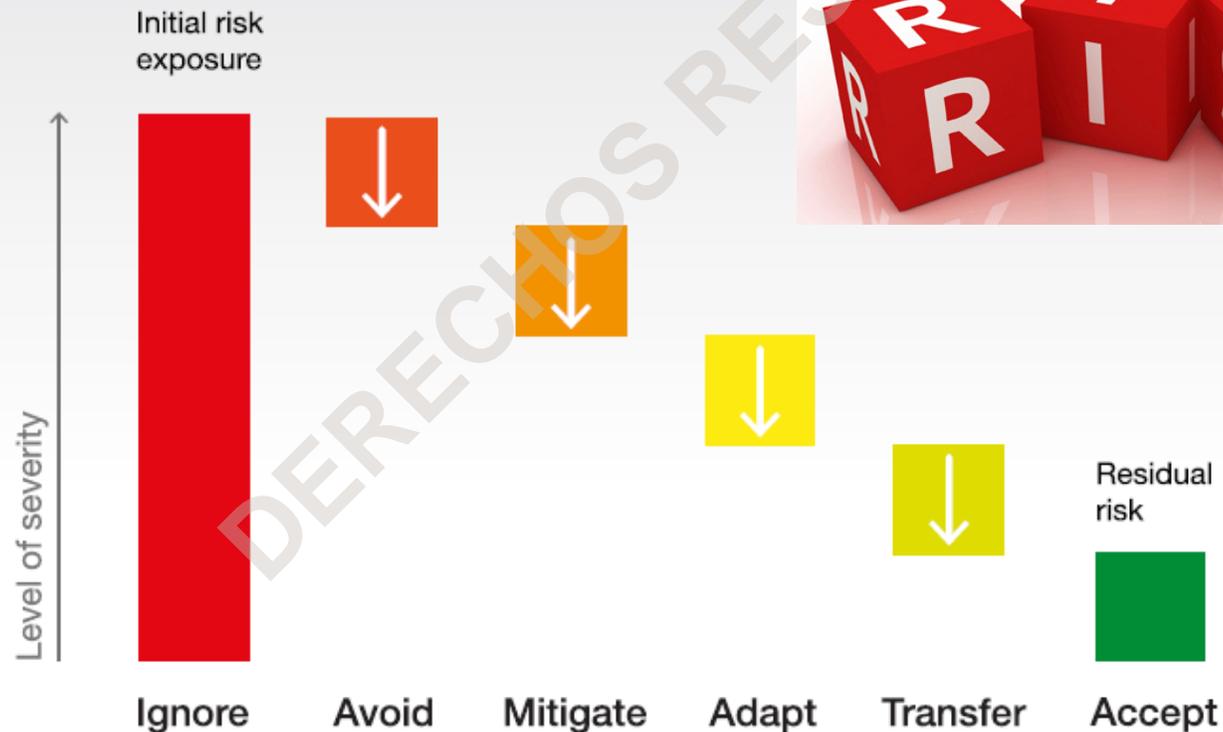
¿Qué les permite decidir sobre las categorías de PREFERIDOS, ACEPTABLES, DESESTIMABLES?

DERECHOS RESERVADOS



# Seleccionar implica Reducir **riesgos** ante proliferación y ante obsolescencias

## Generic risk response strategies

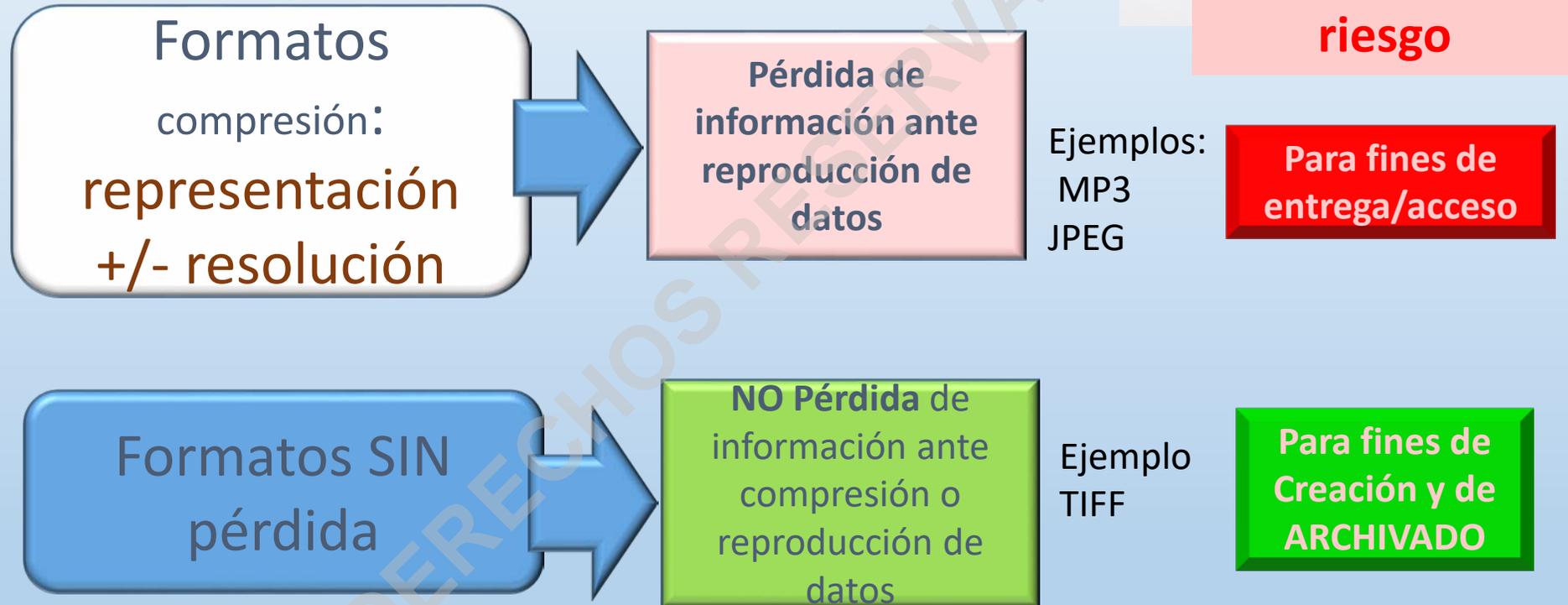


# Un ejemplo

Formatos abiertos	Propietario abierto	Propietario cerrado o puro
Propiedad intelectual públicamente compartida, generalmente mantenida por una organización de estándares.	Propiedad privada pero con propiedad intelectual compartida.	(con derechos de) Propiedad intelectual de propiedad privada 
<b>Riesgo bajo</b>	<b>Riesgo moderado</b>	<b>Riesgo elevado</b>

No todos los ficheros están disponibles en formatos abiertos.  
Ejemplo, realidad virtual, juegos, etc.

# Ejemplo 2º



# Reglas de oro



Seleccionar significa  
actuar con método y rigor

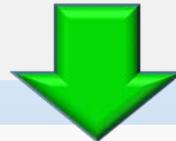
1 Selección de formatos sin pérdida, siempre, para creación y almacenamiento: original y copias de seguridad almacenamiento (fijeza o redundancias).

2 Selección de formatos con pérdida (de representación y resolución) para entrega/acceso temporal.

3 No emplear formatos con pérdida en los ficheros para almacenamiento a largo plazo

# Herramientas para perfilar y evaluar formatos básico de BB Harvard

<https://wiki.harvard.edu/confluence/display/digitalpreservation/Assessment+Tools>



Matriz de formatos Harvard BB accesible en:  
[https://docs.google.com/spreadsheets/d/1buM2XZtkc09kUtUo0W5s0lt4IK\\_6LALF6VooCJDdQZ0/edit#gid=1603236](https://docs.google.com/spreadsheets/d/1buM2XZtkc09kUtUo0W5s0lt4IK_6LALF6VooCJDdQZ0/edit#gid=1603236)

**Criterios de la matriz de selección**

sostenibilidad

costes

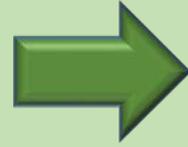
Calidad y funcionalidad

Factores de adopción, de información, de transparencia, de auto documentación, dependencias externas, etc.

Library of Congress: factores  
<https://www.loc.gov/preservation/digital/formats/sustain/sustain.shtml>

Relación de factores	Criterios
----------------------	-----------

<b>Sostenibilidad del formato</b>	Adopción. <b>Divulgación</b> (conocimiento del formato). <b>Transparencia.</b> <b>Auto-documentación.</b> Dependencias externas. <b>Impacto sobre patentes.</b> Soporte para mecanismos de protección.
-----------------------------------	--



<b>Costes</b>	Mantenimiento del entorno para el acceso. Requisitos para la descompresión /almacenamiento Requisitos de almacenamiento con máxima reducción
---------------	--

<b>Funcionalidad y calidad</b>	Apoyo cadena de color. Máximo número de bytes por cadena de color. Resolución Máxima. Capacidad para codificar en modo compresión (color). Ratio de compresión (reducción). Sub-muestreo de croma máxima. Máxima profundidad de bytes en general.
--------------------------------	---

<b>Otros:</b>	Empaquetamiento. Apariencia de la imagen.
---------------	--

Criterios	Indicadores
<b>Divulgación (conocimiento)</b>	<ul style="list-style-type: none"> <li>• Disponibilidad de especificación.</li> <li>• Grado en el que se encuentra completa y es comprensible.</li> <li>• Normalización.</li> </ul>
<b>Impacto de patentes</b>	<ul style="list-style-type: none"> <li>• Restricciones legales que afectan al nuevo uso o a largo plazo.</li> </ul>
<b>Auto-documentación</b>	<ul style="list-style-type: none"> <li>• Tolerancia de errores.</li> <li>• Metadatos técnicos.</li> <li>• Metadatos descriptivos.</li> </ul>
<b>Transparencia</b>	<ul style="list-style-type: none"> <li>• Nivel de complejidad del formato.</li> <li>• Facilidad de identificación exacta.</li> <li>• Grado en el que la compresión es entendida.</li> <li>• Facilidad de validación.</li> <li>• Etc.</li> </ul>

INFORMACIÓN VALIOSA para la *selecting-file-formats*

<http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>

preservación

Adrian Brown, *Head of Digital Preservation Research*  
*Digital Preservation Guidance Note 1: Selecting file formats for long-term preservation*, ANGB, 2008

Relación de **critérios** a considerar para la **preservación a largo plazo**

T  
o  
l  
e  
r  
a  
n  
c  
i  
n  
t  
e  
r  
o  
p  
e  
r  
a  
b  
i  
l  
i  
d  
a  
d

**Ubicuidad**

**Viabilidad**

**Calidad de la documentación**

**Consistencia de los metadatos**

Derechos de propiedad intelectual

**Divulgación de la información**

identificación y validación sencilla

**Complejidad**

E  
s  
t  
a  
b  
i  
l  
i  
d  
a  
d

identificación y validación sencilla

**ejemplo**

<b>Relación de factores</b>	<b>Criterios</b>
<b>Capacidad</b>	Obligada: disponibilidad en diferentes sistemas, búsquedas, transformación a Pdf, apoyo de gráficos, etc. Deseable.
<b>Calidad</b>	<ul style="list-style-type: none"><li>• Precisión.</li><li>• Pérdida.</li></ul>
<b>Resiliencia</b>	<ul style="list-style-type: none"><li>• Ubicuidad.</li><li>• Estabilidad.</li><li>• Recuperación.</li></ul>
<b>Flexibilidad</b>	<ul style="list-style-type: none"><li>• Interoperabilidad.</li><li>• Aplicabilidad.</li></ul>

**Relación de criterios ante la continuidad digital. Fuente: PRO.2011**



## OPF- formato de empaquetamiento



### OPF File Format Risk Registry

Added by [Andrew Jackson](#), last edited by [Johan van der Knijff](#) on Mar 03, 2014 ([view change](#))

- ✓ The OPF File Format Risk Registry provides a reference guide to file format-specific risks that have implications. To add a new Format Registry entry, [follow this link](#). Don't forget to add labels to your new Format Registry entry by clicking on the "Edit" text next to "Labels" at the

Incluir información sobre un formato

Creating page from template [Format Template](#) (choose another template)

Rich Text Wiki Markup Preview

Paragraph **B** *I* U ABC

#### Summary

Name	Format Name
Description	Brief text description of the format
MIME Type(s)	<a href="#">application/octet-stream</a>
PRONOM ID(s)	<a href="#">fmt/xxx</a>
UDFR ID(s)	<a href="#">u1fxxx</a>
Archive Team Wiki	<a href="#">xxx</a>
Library of Congress Digital Formats	<a href="#">xxx</a>
Wikipedia page(s)	<a href="http://wikipedia.org">http://wikipedia.org</a>
File extension(s)	xx.xy
Format specification	<a href="#">^localCopyOfFormatSpecs</a> (Source: <a href="http://www.standardsbody/formatSpecs">http://www.standardsbody/formatSpecs</a> )

#### File Format Risks

- FoxPro
- JP2
  - Effects of bit and byte corruption
  - Files identified as JP2 are really JPX
  - Handling of ICC profiles
  - Image does not comply to technical profile
  - JPEG 2000 support discontinued in Photoshop Elements
  - Lack of performant open source decoding software
  - Not valid JP2
  - Resolution not in expected header fields
- JPX
  - Erroneous tile-part information in images created by Adobe Photoshop
  - Limited viewer and reader support of JPX
- OpenDocument Text
  - Images not embedded because of paste as link bug in OpenOffice and LibreOffice
- Portable Document Format
  - Analysis of Acrobat Engineering PDFs with Acrobat Preflight and Apache Preflight
  - Encryption
  - File attachments
  - Fonts missing, damaged or incomplete

Enlace a riesgos: <http://wiki.opf-labs.org/display/TR/OPF+File+Format+Risk+Registry>



1 -Contextualización del componente “formato” en el documento electrónico.

2 -Retos y medidas adoptadas en la identificación y selección de formatos

3 -Herramientas con funcionalidades diversas para el control de formatos.

4 -Iniciativas *in progres*.

Preservación del documento electrónico: estrategias de identificación y control de formatos.



# Herramientas



Para el control de los formatos

Herramientas para el  
**reconocimiento** de formatos

Herramientas para la  
**identificación** de formatos

Herramientas para la  
**validación** de formatos:  
consistencia fichero y  
metadatos

Herramientas para la  
**Caracterización** de formatos:  
propiedades OD

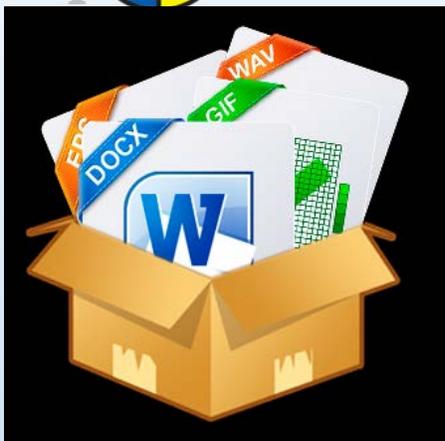
Herramientas para la  
**normalización** de formatos:  
conversión

# Aproximación a cada fichero mediante el análisis de su estructura

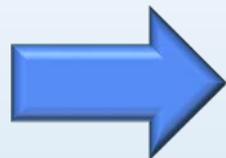
TA2020

Diciembre  
10-11

Semana de Innovación  
Tecnológica en Archivos



*formato*



**GIF87a**

art9.gif	3 KB	Imagen GIF
art10.gif	3 KB	Imagen GIF
layout.lay	4 KB	Archivo LAY
main2.jpg	170 KB	Imagen JPEG
main2.png	208 KB	Imagen PNG
main.jpg	252 KB	Imagen JPEG
main.png	159 KB	Imagen PNG
message.jpg	173 KB	Imagen JPEG
message.png	151 KB	Imagen PNG
options.jpg	215 KB	Imagen JPEG
options.png	151 KB	Imagen PNG

Herramienta para el  
**reconocimiento de  
formatos**

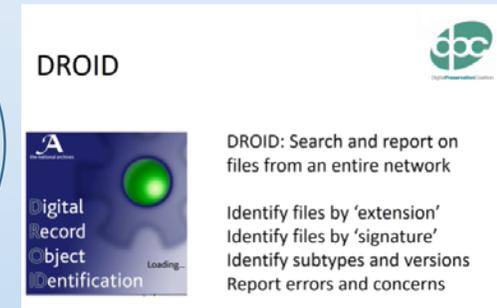
- **<?XML**
- **Doctype svg**
- **Metadatos  
externos**

1. **Extensiones de los ficheros**
2. **Número mágico (*magic number*): extensión y versión**

Formatos modernos

Formatos antiguos

Herramientas para la identificación de formatos



<https://es.wikipedia.org/wiki/TIFF>

La función principal de **DROID** es la **identificación exacta del formato de archivo**, incluso si la extensión de archivo es incorrecta o falta. Aporta información amplia, hasta la versión.

Dispone de **1.353 formatos reconocidos** y se encuentra en incremento constante.

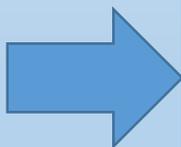
DROID- <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>



## RELACIÓN DE METADATOS EXTRACTADOS POR DROID DE LOS FICHEROS

20  
Diciembre  
10-11

1. ID
2. ID padre
3. identificador único de recursos (URI)
4. file path (columna de recursos en la GUI)
5. **nombre de archivo**
6. **método de identificación** (firma, firma o extensión del contenedor)
7. Estado DROID: guía del usuario
8. tamaño del archivo (columna Tamaño en la GUI)
9. tipo (archivo, carpeta o contenedor)
10. **extensión** de archivo (columna de extensión en la GUI)
11. última fecha de modificación (Última columna modificada en la GUI)
12. Advertencia de **incompatibilidad** de extensión
13. hash (columna Hash en la GUI)
14. **Conteo de formato** de archivo (columna IDs en la GUI)
15. **Identificador único de PRONOM para el formato de archivo (columna PUID en la GUI)**
16. **MIME-type** (columna de tipo MIME en la GUI)
17. **nombre del formato** de archivo (columna Formato en la GUI)
18. **Versión** de formato de archivo (columna Versión en la GUI)





ARCHIVO  
GENERAL  
DE LA NACIÓN  
COLOMBIA

Para convertir

Para  
extensión

categorías

[www.file-extensions.org](http://www.file-extensions.org)

File-Extensions.org  
The Source for File Extensions Information

Find converter Forum Apps Articles FAQ Submit extension

Type file extension Search

**Abrir archivo XLS**  
Libro de Microsoft Excel 97 a 2003

Browse by type

- Extensiones de archivo comunes
- Vídeo
- Audio
- Mapa de bits
- Cámara digital RAW
- Gráficos vectoriales
- Gráficos
- Gráficos 3D, CAD
- Fuente
- Documento
- Texto simple
- Libro electrónico
- Hoja de cálculo
- Microsoft Office relacionado
- Juego de ordenador
- Emulación de software
- Virtualización
- Relacionado con internet
- Relacionado al correo electrónico
- Archivo adjunto de correo electrónico bloqueado
- Peligroso y malicioso
- Cifrado ransomware
- Archivo y comprimido
- Apoyo
- Imagen de disco
- Plataformas móviles
- Tono de llamada móvil
- Archivos contables, financieros,

Impresora Multifunción 0€  
Impresora, Fotocopia, Fax, Escáner. Infórmese sin Compromiso. Tanto Imprimes Tanto Pagas. **ABRIR**

Aprende Idiomas Online  
UOC Xtended Studies  
Aprende Inglés, Francés, Alemán, Chino, Japonés, Catalán y Español **ABRIR**

Una extensión de archivo xls se utiliza tradicionalmente para hojas de cálculo de Microsoft Excel. Una hoja de cálculo es básicamente una cuadrícula de celdas ordenadas en filas y columnas. Cada celda puede contener datos de texto, numéricos o de fórmula. Las hojas de cálculo Excel XLS se pueden abrir en la mayoría de los otros programas similares, como OpenOffice.org Calc.

Descripción detallada  
Software que abre el archivo xls  
Agregar comentario

Marcar y compartir esta página con otras personas:  
f 8 183 t +

**Extensión de archivo XLS : libro de Microsoft Excel 97 a 2003**

¿Qué es el archivo xls? ¿Cómo abrir archivos xls?

La extensión de archivo xls está asociada con Microsoft Excel, una de las herramientas más populares y poderosas que puede utilizar para crear hojas de cálculo de formato, gráficos y mucho más.

Los archivos xls se utilizan en Microsoft Excel 97 a 2003 para libros de trabajo, archivos de documentos de hojas de cálculo. El XLS es el formato de archivo propietario de Microsoft Excel para almacenar sus documentos.

Microsoft Excel hasta la versión 2007 utilizó un formato de archivo binario propietario llamado Formato de archivo de intercambio binario (BIFF) como su formato principal. Hay varias versiones del formato que se han publicado a medida que fue revisado a lo largo del tiempo.

Especificación de tipo de archivo:  
Tipo de archivo de Microsoft Office  
Tipo de archivo de hoja de cálculo y libro de trabajo  
icono de archivo xls: XLS

Artículos relacionados:  
Cómo abrir documentos de Microsoft Office y OpenOffice  
Abrir y convertir hojas de cálculo de números en línea  
Los mejores convertidores en línea  
Extraer contactos de archivos SPB  
Recuperar documentos no guardados o corruptos  
Convertir archivos dentro de Google Drive  
Tipos de archivos soportados en Microsoft Excel 2010  
Cómo convertir documentos de Excel a otro formato de archivo  
Cómo cambiar la carpeta predeterminada de documentos de Windows

Ayuda como abrir:  
Puede abrir los \*.xls en Microsoft Excel o cualquier otra aplicación que pueda manejar el formato, como OpenOffice.org,

Información  
sobre el  
formato

IDENTIFICA  
CARACTERIZA Y  
NORMALIZA

BITA2020  
Diciembre  
10-11  
ana de Innovación  
ológica en Archivos



ARCHIVO  
GENERAL  
DE LA NACIÓN  
COLOMBIA



Dato  
curioso

Database files	file extension <b>ktm</b>	Unknown file extension
GIS, GPS navigation and map files	file extension <b>flex</b>	Unknown file extension
Program executable files	file extension <b>jpj</b>	Unknown file
System files	file extension <b>kik</b>	Unknown file extension
Settings, options, themes or skins files	file extension <b>pwc7</b>	Unknown file extension
Plugin, addon files	file extension <b>wwz</b>	Unknown extension
Source code and script files	file extension <b>hug</b>	Unknown file extension
Configuration files	file extension <b>pthc</b>	Unknown file extension
Binary files	file extension <b>6_53_</b>	Unknown or unassigned file
Encoded and encrypted files	file extension <b>ziw</b>	Unknown file extension
Temporary files	file extension <b>fgw</b>	Unknown file extension
Unknown file types	file extension <b>bkh</b>	Unknown file extension
Various data files	file extension <b>x2d</b>	Unknown file extension
Misc file extensions	file extension <b>noname</b>	Unknown file extension
	file extension <b>tdlex</b>	Unknown file extension
	file extension <b>pubx</b>	Unknown file extension
	file extension <b>tbg</b>	Unknown file extension
	file extension <b>failed-conv</b>	Unknown file extension
	file extension <b>hxl</b>	Unknown file extension
	file extension <b>yesterday</b>	Unknown file
	file extension <b>pwbk</b>	Unknown file extension
	file extension <b>zfvvv</b>	Unknown file extension
	file extension <b>wh</b>	Unknown file extension
	file extension <b>pushstore</b>	Unknown file extension

# SITA2020

Semana de Innovación  
Tecnológica en Archivos

Diciembre  
10-11

INCLUYE una  
relación de  
formatos  
desconocidos y  
petición de  
información



**categorias**

**JNLP**

**File Type** Java Web Start File

**Developer** Oracle ← **creador**

**Popularity** ★★★★★

**Category** Web Files ← **Tipo o categoría de recurso**

File created in the Java Network Launching Protocol (JNLP) format; used for launching and managing Java programs over a network or on the Web (e.g., an applet in a Web browser); can be double-clicked to run the program if the Java Runtime Environment (JRE) is installed.

**File Format Description**

JNLP files contain information such as the remote address for downloading the Java program (JAR file) and the initial class to run. They are saved in an XML format and can be viewed or edited with a text editor.

**Descripción**

**Program(s) that open .JNLP files**

Windows

- Oracle Java Web Start
- Any text editor

Mac OS

- Oracle Java Web Start
- Any text editor

extensión

Descripción

Aplicaciones que lo abren

**Funcionalidad de FILEINFO.NET**

**Identificar** el formato de fichero

**Caracterizar:** Describe propiedades



## Library of Congress. 2004

Diciembre  
10-11

### Sustainability of Digital Formats Planning for Library of Congress Collections

[NDIIPP Home](#) | [Digital Formats Home](#)

[Introduction](#) | [Sustainability Factors](#) | [Content Categories](#) | [Format Descriptions](#) | [Contact](#)

De 2004 hasta 2015.  
De 2015-Iniciativa  
Library of Congress

**Introduction** >> [Overview](#) | [Formats, Evaluation Factors, and Relationships](#) | [Papers and Presentations](#) | [Related Resources](#)

#### Overview

##### Table of Contents

- [Digital formats: a work in progress](#)
- [Purposes](#)
- [Scope](#)
- [Typical questions the site will answer](#)
- [More information](#)

#### Digital formats: a work in progress

The Digital Formats Web site provides information about digital content formats. An initial offering was placed online in 2004 and expanded and updated analyses and resources have been added regularly. Digital formats will continue to evolve in the coming years and this or a successor site will also evolve to keep pace.

[https://www.loc.gov/preservation/digital/formats/content/content\\_categories.shtml](https://www.loc.gov/preservation/digital/formats/content/content_categories.shtml)

### Sustainability of Digital Formats Planning for Library of Congress Collections

[Introduction](#) | [Sustainability Factors](#) | [Content Categories](#) | [Format Descriptions](#) | [Contact](#)

**Content Categories** >> [Still Image](#) | [Sound](#) | [Textual](#) | [Moving Image](#) | [Web Archive](#) | [Datasets](#) | [Geospatial](#) | [Generic](#)

#### Content Categories

This offering began with descriptions of still images, sound recordings, textual content, and moving images, a starting point that reflected the Library's experience with the types of digital reproductions of historical content produced for [American Memory](#), and general familiarity with its physical collections of photographs, pictorial prints, sound recordings, books and periodicals, and moving image items. These four basic content categories also serve as building blocks for more complex digital works, e.g., serials, compilations, Web sites, multimedia works, etc. Subsequent additions have included the formats appropriate for archived Web sites (the Library of Congress has an active [Web Archiving project](#)), datasets (focusing on sets of data values in an organized structure intended for automated analysis), and geospatial information. Over time, additional descriptions will be added to all categories and new categories will be added as well.

Note that descriptions also exist for [Generic Formats](#).



[Content Categories](#) >> [Still Image](#) | [Sound](#) | [Textual](#) | [Moving Image](#) | [Web Archive](#) | [Datasets](#) | [Geospatial](#) | [Generic](#)

### Textual Content

- [Quality and functionality factors](#)

Factors appropriate for the evaluation of digital formats for text.

- [Preferences in summary](#)

Abbreviated overview of formats for text preferred by the Library of Congress.

- [Curator's view](#)

Illustrative view of a curator's identification of significant characteristics of textual content and preferred formats.

- [Format description documents](#)

More information about specific digital formats for text.

[Format Description Categories](#) >> [Browse Alphabetical List](#)

### Format Descriptions for Text

[>> Back](#)

- [DTB](#)
  - [DTB\\_2002](#), DTB (Digital Talking Book), 2002
  - [DTB\\_2002\\_NLS](#), DTB (Digital Talking Book), NLS Extension for AMR-WB+ Speech Codec
  - [DTB\\_2005](#), DTB (Digital Talking Book), 2005
- [EPUB](#)
  - [EPUB\\_2](#), EPUB, Electronic Publication, Version 2
  - [EPUB\\_3](#), EPUB, Electronic Publication, Version 3
- [OEBPS](#)
  - [OEBPS\\_1\\_0](#), Open eBook Forum Publication Structure 1.0.1
  - [OEBPS\\_1\\_2](#), Open eBook Forum Publication Structure 1.2
- [NCBI/NLM](#)
  - [NCBIArch\\_1](#), NCBI/NLM Journal Archiving and Interchange DTD, version 1
- [NewsML\\_1](#)
  - [NewsML\\_1](#), NewsML (News Markup Language), version 1
- [NITF](#)
  - [NITE](#), News Industry Text Format
- [PDF](#)

General notes about textual formats are provided by [Text Quality and Functionality Factors](#).

The descriptions listed on this page provide specific information about individual file formats, file-format classes, bitstream structures and encodings, and the mechanisms used to compress files or bitstreams. Inclusion of a format does not imply that it is preferred or acceptable for Library of Congress collections. Conversely, omission of a format from the list does not imply that it is not preferred or acceptable. Descriptions will be drafted and added over time. Persons with specialized knowledge are encouraged to review and comment on the descriptions using this [contact form](#).



ARCHIVO  
GENERAL  
DE LA NACIÓN  
COLOMBIANA

## Sustainability of Digital Formats Planning for Library of Congress Collections

[Introduction](#) | [Sustainability Factors](#) | [Content Categories](#) | [Format Descriptions](#) | [Contact](#)

[Format Description Categories](#) >> [Browse Alphabetical List](#)

### Format Descriptions for Text

[>> Back](#)

- *DTB*
  - [DTB\\_2002](#), DTB (Digital Talking Book), 2002
  - [DTB\\_2002\\_NLS](#), DTB (Digital Talking Book), NLS Extension for AMR-WB+ Speech Codec
  - [DTB\\_2005](#), DTB (Digital Talking Book), 2005
- *EPUB*
  - [EPUB\\_2](#), EPUB, Electronic Publication, Version 2
  - [EPUB\\_3](#), EPUB, Electronic Publication, Version 3
- *OEBPS*
  - [OEBPS\\_1\\_0](#), Open eBook Forum 1.0.1
  - [OEBPS\\_1\\_2](#), Open eBook Forum 1.2
- *NCBI/NLM*
  - [NCBIArch\\_1](#), NCBI/NLM Journal Article DTD, version 1
- *NewsML\_1*
  - [NewsML\\_1](#), NewsML (News Marking Language), version 1
- *NITF*
  - [NITF](#), News Industry Text Format, version 1
- *PDF*

General notes about textual formats are provided by [Text Quality and Functionality Factors](#).

The descriptions listed on this page provide specific information about individual file formats, file-format classes, bitstream structures and encodings, and the mechanisms used to compress files or bitstreams. Inclusion of a format does not imply that it is preferred or acceptable for Library of Congress collections. Conversely, omission of a format from the list does not imply that it is not preferred or acceptable. Descriptions will be drafted and added over time. Persons with specialized

## Sustainability of Digital Formats Planning for Library of Congress Collections

[Introduction](#) | [Sustainability Factors](#) | [Content Categories](#) | [Format Descriptions](#) | [Contact](#)

[Format Description Categories](#) >> [Browse Alphabetical List](#)

### EPUB, Electronic Publication, Version 2

[>> Back](#)

#### Table of Contents

- [Identification and description](#)
- [Local use](#)
- [Sustainability factors](#)
- [Quality and functionality factors](#)
- [File type signifiers](#)
- [Notes](#)
- [Format specifications](#)
- [Useful references](#)

# SITA2020

Semana de Innovación  
Tecnológica en Archivos

Diciembre  
10-11

## PROPIEDADES

#### Format Description Properties

- ID: fdd000278
- Short name: EPUB\_2
- Content categories: text
- Format Category: encoding, file-format
- Other facets: container-wrapper, text, structured, symbolic
- Last significant update: 2011-03-31
- Draft status: Full



#### Identification and description

Full name	EPUB, Electronic Publication, Version 2
Description	<p>EPUB is a format for electronic publications with reflowable text in marked up document structure with associated images for illustrations, all in a container format. Reflowable text allows the text display to be optimized for the particular display device used by the reader of the EPUB-formatted book; this is in contrast to documents with pre-determined pagination. EPUB allows publishers to control document presentation through style-sheets. EPUB, Version 2 comprises three separate specifications:</p> <ul style="list-style-type: none"> <li>• OCF, Open Container Format, specifies the mandatory container for an EPUB. OCF is based on the ZIP format.</li> <li>• OPS, Open Publication Structure specifies options for representation of the content of an electronic publication. Content components can be XML-based documents for the flowable text, together with raster or vector images representing illustrations. OPS specifies "preferred vocabularies" for XML-based content components and "core</li> </ul>



## Sustainability of Digital Formats Planning for Library of Congress Collections

[Introduction](#) | [Sustainability Factors](#) | [Content Categories](#) | [Format Descriptions](#) | [Contact](#)

**Introduction** >> [Overview](#) | [Formats, Evaluation Factors, and Relationships](#) | [Papers and Presentations](#) | [Related Resources](#)

## Sustainability of Digital Formats Planning for Library of Congress Collections

[Introduction](#) | [Sustainability Factors](#) | [Content Categories](#) | [Format Descriptions](#) | [Contact](#)

**Content Categories** >> [Still Image](#) | [Sound](#) | [Textual](#) | [Moving Image](#) | [Web Archive](#) | [Datasets](#) | [Geospatial](#) | [Generic](#)

### Textual Content

- [Quality and functionality factors](#)

Factors appropriate for the evaluation of digital formats for text.

- [Preferences in summary](#)

Abbreviated overview of formats for text preferred by the Library of Congress.

- [Curator's view](#)

Illustrative view of a curator's identification of significant characteristics of textual content and preferred formats.

- [Format description documents](#)

More information about specific digital formats

formats. An initial  
and resources have  
g years and this or

Diciembre  
10-11

### Funcionalidad de sostenibilidad

**Identificar** el formato de fichero, si dispone de él en su base

**Caracterizar:** datos técnicos ORIENTATIVOS para conocer las propiedades del objeto, pero sobre todo cara a la **CONSERVACIÓN** de los formatos.



<http://xena.sourceforge.net/>

## Dos tareas

El software Xena ayuda a la preservación digital realizando dos tareas importantes:

- **Detectar los formatos** de archivo de objetos digitales
- **Convertir objetos** digitales en **formatos abiertos** para la preservación *Normalising for Archives*.

←  
identificador

←  
normalizador

## Funcionalidad de Xena

**Identificar** el formato de fichero, si dispone de él en su base

**Normalizar** Convierte el fichero en un formato para la preservación en código abierto.



ARCHIVO  
GENERAL  
DE LA NACIÓN  
COLOMBIA

<http://jhove.sourceforge.net/index.html>

**SITA2020**

Semana de Innovación  
Tecnológica en Archivos

Diciembre  
10-11

**JH**  **OVE**

**JH**  **VE2**

2002-

HOME TUTORIAL DOCUMENTATION DISTRIBUTION LINKS MAILING LIST JHOVE PROJECT

**JHOVE - JSTOR/Harvard Object Validation Environment**

1 Introduction

En general, los procesos que desempeña esta aplicación responden a cuatro preguntas sobre un objeto digital:

- ¿Qué es?
- ¿Qué es, realmente o qué tipo de propiedades presenta?
- ¿Qué pasa con su comportamiento?
- ¿Y qué tal de aceptable es?

identificación, validación,  
extracción de características y  
evaluación.

Facilita la captura , gestión y almacenamiento de un PIT a PIA

Biblioteca de la  
Universidad de Harvard



Identificación del formato  
de archivo



Considerando *su extensión y su número mágico*.



Validación



determina el *nivel de conformidad de un objeto digital mediante reglas normativas y sintácticas* definidas por la especificación.



Extracción de  
características



Informe sobre las *propiedades intrínsecas de un objeto digital significativas* para la preservación.



La evaluación



Determinar el *nivel de aceptabilidad* de un objeto digital para un propósito específico sobre la base de unas políticas definidas.

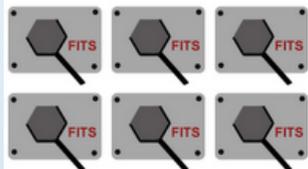
# funcionalidades

# FITS → IDENTIFICA, VALIDA Y EXTRACTA METADATOS TÉCNICOS PARA FORMATOS DE FICHEROS

ITA2020

na de Innovación  
lógica en Archivos

Diciembre  
10-11



Conjunto de herramientas de  
información de archivo  
(FITS)

Documentación y lanzamiento de código  
oficial de los proyectos FITS y FITS Web  
Service

[Contacto](#)



University of Harvard

DESCARGAS DOCUMENTACIÓN NOTICIAS PRESENTACIONES

Guía de inicio rápido

Manual de usuario

Manual del desarrollador

Grupo de discusión

Preguntas más frecuentes

Cuestiones

Proyecto en casa

[INICIO /](#)

## Descargas

- [ENCAJA](#)
- [Servicio web FITS](#)

**FITS** : las versiones oficiales de FITS están disponibles aquí en formato ZIP. Para  
la versión o descargar el código fuente, visite [nuestra página de github de FITS](#)  
es proporcionada por un enlace estático en la parte inferior de la tabla como fit

Nombre del	Resumen	Fecha de	Talla
------------	---------	----------	-------

<http://projects.iq.harvard.edu/fits>

<http://projects.iq.harvard.edu/fits/downloads>

## Funcionalidad de FITS

**Identificar el formato de fichero**  
**Validar y extraer metadatos**





## The technical registry PRONOM

[Welcome](#) : [About](#) [Add an entry](#)  
[Search](#) ? [Help](#) [Information resources](#)

? [Help](#) : search by format

 Search : By format

[Simple search](#) [File format](#) [PRONOM Unique Identifier](#) [Software](#) [Vendor](#) [Lifecycles](#) [Migration Pathways](#)

### 1. File formats

Enter a file extension and click 'search' to find all file formats with that extension. Leave the file extension blank to find all file formats in the database.

\*.  [Search >](#)

To search for a particular file format, enter the name of the file format and then click 'search'. Leave the file format name blank to find all file formats in the database.

[Search >](#)

### 2. Compatible software

Enter a file extension and click 'search' to find  software which can  files with that extension.

\*.  [Search >](#)

Enter a file format name and click 'search' to find  software which can  files of that format.

[Search >](#)

### 3. File format risk

Click Search to find  formats with a risk   [Search >](#)

Numerosa  
información  
para cada  
formato o  
extensión de  
formato



Tools	Identifica	Valida	Caracteriza	Normaliza
File- extensión				
Fileinfo.net				
Droid			Con PRONOM	
Xena		metadatos		
FITS		metadatos		
Apache Tika		metadatos		
Jhove2		metadatos		
PRONOM		metadatos		

Comparativa de funcionalidad de aplicaciones de los formatos



ARCHIVO  
GENERAL  
DE LA NACIÓN  
COLOMBIA

# Herramientas para identificación de formatos

Más

**COPTR**



Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark



# SITA2020

Semana de Innovación  
Tecnológica en Archivos

Diciembre  
10-11

[http://coptr.digipres.org/Category:File Format Identification](http://coptr.digipres.org/Category:File%20Format%20Identification)



## Categoría: Identificación del formato de archivo

### Páginas de la categoría "Identificación del formato de archivo"

Las siguientes 23 páginas pertenecen a esta categoría, de un total de 23.

#### UNA

- Apache Tika

#### re

- DROID (Identificación de objetos de registro digital)
- Utilidad DUMPBIN

#### mi

- EssArch

#### F

- FIDO (identificación de formato para objetos digitales)
- FIDOO
- Comando de archivo libre fino
- FITS (conjunto de herramientas de información de archivos)

#### GRAMO

- Gvfs-info

#### J

- JHOVE (entorno de validación de objetos de Harvard)
- JHOVE2

#### L

- Libmagic-dev
- Libsharedmime

#### METRO

- MediaConch

#### norte

- Nanita
- Analizador de archivos NARA y recolector de metadatos

#### O

- Officeparser.py
- Oh cuenta

#### PAG

- Utilidad de desarrollo de firmas PRONOM

#### S

- Siegfried

#### T

- Identificador de archivo TrID

#### V

- Vitam

#### W

- Descubrimiento de archivos web

**Definición de función:** herramientas que permiten la identificación automática del formato de archivo de un archivo en particular, por lo general mediante el examen de los códigos de características (a menudo denominado formato de archivo mágico) en el encabezado del archivo.

Categorías : Función | Ingerir

<https://coptr.digipres.org/Category:File Format Identification>



## Categoría: Validación

### Páginas de la categoría "Validación"

Las siguientes 39 páginas pertenecen a esta categoría, de un total de 39.

#### 3

- Validador de PDF de 3 alturas (TM)

#### UNA

- Prueba ADC
- Apache PDFBox

#### segundo

- Mala peggy
- BitCurator
- MetaEdit BWF

#### C

- Checkit tiff
- Validador CSV

#### re

- Gerente de DPF

#### mi

- EpubCheck
- EssArch

#### F

- Analizador de archivos y recopilador de metadatos V2
- FITS (conjunto de herramientas de información de archivos)

#### F cont.

- Pederaln

#### yo

- ImageVerifier

#### J

- JHOVE (entorno de validación de objetos de Harvard)
- JHOVE2
- Jp2StructCheck
- Jpylyzer
- JWAT

#### K

- KOST-Simy
- KOST-Val

#### METRO

- MediaConch
- MP3val

#### norte

- Namalsator

#### O

- Validador ODF

**Definición de función:** herramientas que admiten la validación de archivos digitales, generalmente contra una especificación de formato de archivo.

#### PAG

- Manejo de paquetes
- Herramientas PDF (por Didier Stevens)
- PdfaPilot
- PDFTron PDF-A Manager
- Herramienta de preingestión
- PREMIS en la caja de herramientas METS (PiM)
- Validador Python DPX

#### S

- SIARD-VAL

#### T

- TIFF-Val

#### V

- Validador (por CCS)
- VeraPDF

#### W

- Servicio de validación de marcado W3C
- Warctools

Las siguientes 73 páginas pertenecen a esta categoría, de un total de 73.

## UNA

- Elementos de Adobe Photoshop
- Apache PDFBox
- Apache POI: la API de Java para documentos de Microsoft
- Apache Tika

## segundo

- BitCurator
- Brunilda
- MetaEdit BWF

## C

- C3PO
- CyberChef

## re

- DIMAG
- DIMAG IngestList
- Lista de directorio e impresión
- DisclmageChef
- Tipo de disco
- DROID (Identificación de objetos de registro digital)
- Motor de análisis DROID Siegfried Sqlite
- Utilidad DUMPBIN

## mi

- Extractor fácil de CD-DA
- EMET (herramienta de extracción de metadatos integrada)
- EPADD
- EpubCheck
- EssArch
- Copia de audio exacta
- Explorador de EXE
- Exempi

## E cont.

- Normalizador EXIF a DC XML
- ExifTool

## F

- FIDO (identificación de formato para objetos digitales)
- FIDOO
- Analizador de archivos y recopilador de metadatos V2
- FileAlyzer
- FITS (conjunto de herramientas de información de archivos)

## GRAMO

- Geosetter
- GetID3 ()
- Libextractor GNU

## yo

- Index.dat Analyzer v2.5
- Yo texteo

## J

- JHOVE (entorno de validación de objetos de Harvard)
- JHOVE2
- Jp2StructCheck
- Jpylyzer
- JWAT

## K

- Frase de Keith Humphreys

## L

- Lingfo

## METRO

- Inicio de MailStore
- Mdqc
- Información de los medios
- Herramienta de extracción de metadatos
- Interrogador de metadatos
- Transformador de metadatos

## M cont.

- Metadatos ++
- MP3 :: Etiqueta

## norte

- Nanita
- Analizador de archivos NARA y recolector de metadatos
- Analizador de fotogramas de vídeo NARA

## O

- Validador ODF
- Officeparser.py
- OpenJPEG

## PAG

- Pagelyzer
- Herramientas PDF (por Didier Stevens)
- PdfaPilot
- Pdftk
- Peepdf
- Herramienta de extracción PERICLES (PET)
- Herramienta de preingestión
- Kit de herramientas de Python XMP

## Q

- Qpdf

## R

- Rata

## V

- VRenamer

## W

- Warctools
- Descubrimiento de archivos web
- WordHoard

## X

- Xpdf



1 -Contextualización del componente “formato” en el documento electrónico.

2 -Retos y medidas adoptadas en la identificación y selección de formatos

3 -Herramientas con funcionalidades diversas para el control de formatos.

4 -Iniciativas *in progres*.

Preservación del  
documento  
electrónico:  
estrategias de  
identificación y  
control de formatos.

[https://wiki.dpconline.org/index.php?title=File Formats Assessments](https://wiki.dpconline.org/index.php?title=File_Formats_Assessments)



Página [Discusión](#)

[Leer](#)

[Ver fuente](#)

[Ver historial](#)

Buscar wiki.dpconline.org



## Evaluaciones de formatos de archivo



Los formatos de archivo son un medio de estructurar la información de manera sensata para el almacenamiento, la recuperación y el uso. Hay una gran cantidad de formatos diferentes que admiten una variedad de tipos de datos, desde instancias específicas hasta formatos de contenedor capaces de almacenar diferentes tipos de datos.

Como se discutió en su documento de IPRES "Evaluaciones de sostenibilidad en la Biblioteca Británica: Formatos, marcos y resultados", el Equipo de Preservación Digital de la Biblioteca Británica ha llevado a cabo evaluaciones de formato de archivo para capturar el conocimiento sobre las brechas en las mejores prácticas actuales, la comprensión y la capacidad de trabajo con formatos de archivo específicos. El enfoque de cada evaluación es capturar los riesgos de preservación basados en evidencia y las implicaciones de la obsolescencia institucional que conducen a problemas para mantener el contenido a lo largo del tiempo.

Las evaluaciones de la Biblioteca Británica están disponibles a través de esta página wiki de DPC para compartir sus hallazgos y facilitar el compromiso con la comunidad más amplia de preservación.

La retroalimentación siempre es bienvenida. Si tiene algún comentario o sugerencia, envíe un correo electrónico a: [DPT en la Biblioteca Británica](mailto:DPT@britishlibrary.gov.uk)

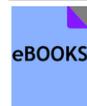


### Colaboración

La Biblioteca Británica, la Biblioteca del Congreso, la Biblioteca de Harvard, NARA y la Digital Preservation Coalition están comenzando una nueva colaboración para coordinar y poner a disposición sus evaluaciones de formato de archivo. Esto aumentará el conjunto de evaluaciones disponibles, evitando la duplicación, aumentando la calidad y minimizando el esfuerzo de mantenimiento. Como primera etapa, estas organizaciones están coordinando su próximo trabajo de evaluación aquí.

- [Evaluaciones de la Biblioteca del Congreso](#)
- [Evaluaciones de la Biblioteca de la Universidad de Harvard](#)
- [Evaluaciones de la Administración de Registros de Archivos Nacionales \(próximamente\)](#)

### Evaluaciones de riesgo de preservación - Resúmenes



#### Resumen de eBook

Una amplia descripción de los formatos disponibles en el sector de libros electrónicos.

### Evaluaciones de riesgos de preservación por tipo de formato

#### Criterios de evaluación

Consulte los [factores de evaluación de formato](#) cubiertos en cada evaluación.

#### FORMATOS DE IMAGEN



#### Formato de archivo de imagen etiquetada

Un formato ráster ampliamente compatible para imágenes.



#### JPEG 2000

Un estándar de compresión y un sistema de codificación para imágenes, creado por el Joint Photographic Experts Group.

# British Library



<b>Digital Preservation Team</b>	<b>Preservation Assessment:</b> PDF Format (Part 1)	<b>Date:</b> 30/06/2019
		<b>Version:</b> 1.5

### **PDF Format Preservation Assessment Part 1: PDF**

#### *Document History*

Date	Version	Author(s)	Circulation
10/12/2014	1.1	Paul Wheatley, Peter May, Maureen Pennock	External
25/02/2015	1.2	Peter May	External
04/09/2015	1.3	Simon Whibley	External
11/07/2017	1.4	David Russo	
30/06/2019	1.5	Peter May, Akiko Kimura. (Minor update for the core PDF with the review by the National Library of Wales)	External (Restructured as a two-part document. See PDF Format Preservation Assessment, Part 2: PDF/A Profile, Version 1.0)

British Library Digital Preservation Team  
[digitalpreservation@bl.uk](mailto:digitalpreservation@bl.uk)

[https://wiki.dpconline.org/images/f/ff/PDF\\_Assessment\\_v1.5.pdf](https://wiki.dpconline.org/images/f/ff/PDF_Assessment_v1.5.pdf)

# SITA2020

Semana de Innovación  
Tecnológica en Archivos

Diciembre  
10-11

# ejemplo



## Formato Fichas

- Estructura de las fichas validée
- Correcciones de imágenes
  - JPEG 2000 BnF validée
  - Formato de intercambio de archivos JPEG (JPEG) à valider
  - PELEA validée
- Contenu textuel
  - Texte numérisé
    - Físico de estructuración (segmentation et reconnaissance des caractères)
      - XML-ALTO validée
    - Structuration intellectuelle (reconocimiento de la mise en page, table des matières...)
      - XML-tdmNum validée
      - XML-METS validée
  - Texte nativement numérique, formats de publication
    - PDF / A validée
    - EPUB validée
    - DTBook (DAISY XML) validée
- Texte nativement numérique, formats de production
  - Abrir texto del documento en cours
- Contenu sonore
  - MPEG 1/2 Audio Layer III (MP3) validée
  - Audio MPEG 2/4 (AAC) en cours
  - FLAC nativo validée
  - ONDA / ONDA RF64 validée
- Imagen animada
  - Apple ProRes (códec) validée
  - DNxHD (códec) en cours

▼ Páginas 25

Encuentra una página...

- Hogar
- CAA
- ALTO
- DNxHD
- DTBook (DAISY XML)
- EPUB
- H.264
- H.265
- Estándar de metadatos de fotografías IPTC
- JFIF\_JPEG
- JPEG\_2000\_BnF
- MBOX
- METS
- MOV
- MP3

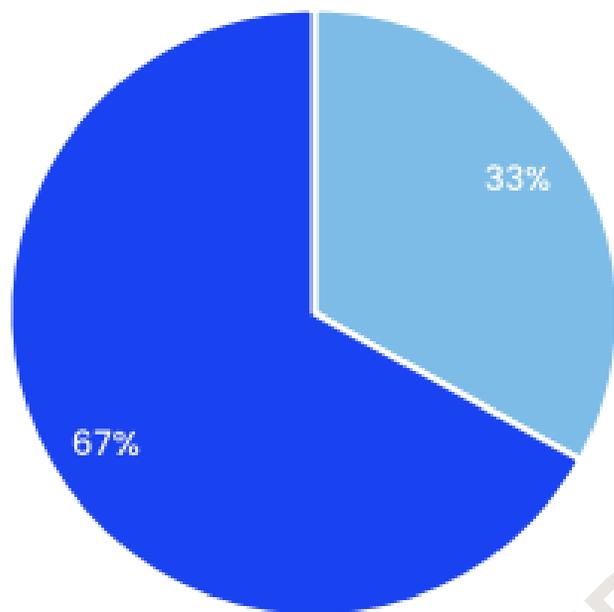
Mostrar 10 páginas más ...

Clonar esta wiki localmente

<https://github.com/hackathonBnF/>



<https://www.netwerkdigitaalrfgood.nl/wp-content/uploads/2020/06/NDE-The-Use-of-Preservation-Tools-among-Dutch-Heritage-Organizations-juni2020-2.pdf>



- Instructions for the use of tools, manuals for the use of tools, interpretation of errors, development of tools, maintenance of tools, other
- Best practice examples



Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Figure 22: Where would you like to collaborate with others in the area of preservation tools?



dutch digital  
heritage  
network



# Necesidades detectadas

Herramientas para la  
identificación de  
formatos de ficheros:  
emails, website, social  
media, OD  
tridimensionales, etc.

Herramientas para la  
extracción de  
metadatos

Herramientas para la  
normalización y  
conversión

Fallos en uso de  
herramientas en procesos  
de pre-ingesta e ingesta:  
redundancias, encriptados,  
análisis de dependencias..



<https://openpreservation.org/news/virtual-research-environment-for-preservation-tools/?order=date%3ADESC&cat=0&page=2>

SITA2020

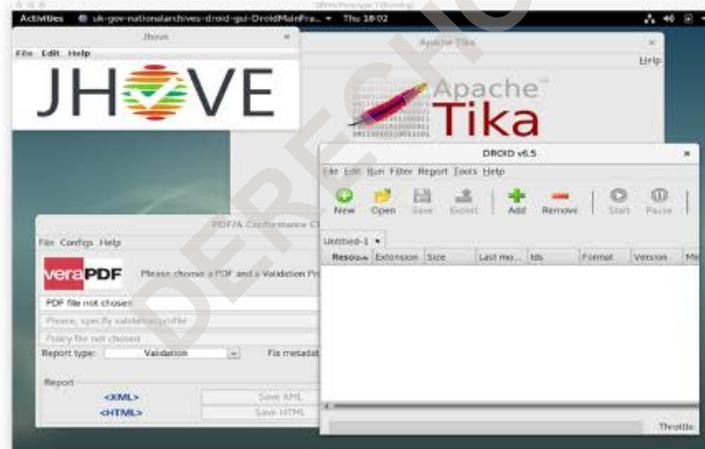
Diciembre  
10-11

## Entorno de investigación virtual para herramientas de conservación

De julio de 21 de , 2020

La Red de Patrimonio Digital Holandés (DDHN) y la Fundación de Preservación Abierta (OPF) están trabajando juntas para desarrollar un Entorno de Investigación Virtual (VRE) con herramientas de preservación. Tanto la [encuesta de la comunidad OPF](#) como una [encuesta realizada por DDHN entre las organizaciones de patrimonio holandesas](#) identificaron una creciente necesidad entre la comunidad internacional de trabajar con herramientas de preservación de código abierto.

Este proyecto tiene como objetivo responder a esta necesidad mediante la creación de una máquina virtual fácil de instalar que ejecute algunas de las herramientas de preservación de código abierto más populares. Las organizaciones patrimoniales podrán utilizar el VRE para probar herramientas de preservación sin tener que instalarlas. La versión actual del entorno de investigación virtual incluye [Apache Tika](#) , [DROID](#) , [JHOVE](#) y [veraPDF](#) . El equipo está probando el primer prototipo y espera compartir más noticias después del verano.



DROID

DDHN + OPF

¿Está interesado en probar el VRE? ¿Hay herramientas que le gustaría ver incluidas? ¿Alguna otra pregunta? Si es así, no dude en ponerse en

# SITA2020

Semana de Innovación  
Tecnológica en Archivos



Diciembre

10-11

# ¡GRACIAS!

[moroca@usal.es](mailto:moroca@usal.es)

DERECHOS RESERVADOS



VNIVERSIDAD  
DSALAMANCA  
CAMPUS DE EXCELENCIA INTERNACIONAL



800 AÑOS  
1218 - 2018