

# Identificación, caracterización y selección de formatos para la preservación del recurso digital

## *Identification, characterization and selection of the formats for the preservation of the digital resource*

**Manuela Moro Cabero**

moroca@usal.es

**Profesora Titular. Universidad de Salamanca**

<https://orcid.org/0000-0001-5301-1924>

### **Resumen**

Los formatos evolucionan a medida que los desarrolladores identifican e incorporan nuevas funciones para las aplicaciones. Se encuentran supeditados a elevada proliferación y obsolescencia, generando al profesional responsable de la preservación de recursos digitales, problemas y obligaciones. El estudio de naturaleza descriptiva, analítica y comparativa, responde al interrogante: ¿Qué se debe conocer y cómo se debe actuar ante esta situación? Facilita una metodología para identificar y gestionar formatos; aporta un estudio comparado sobre las prestaciones observadas en las herramientas de identificación y control.

### **Palabras clave**

Formatos; Preservación digital; Herramientas

### **Abstract**

The formats evolve as the developers identify and incorporate new purposes for the applications. They can be found dependent on a high proliferation and obsolescence, producing problems and obligations to the professional responsible of the preservation of digital resources. the study, descriptive, analytical and comparative in nature, answers the question: What must be known and how this situation must be faced? It provides a methodology to identify and manage the formats, with a comparative study about the observed outputs in the identification and control tools.

### **Keywords**

File Formats; Digital preservation; Tools

Recibido: 11/05/2018

Aceptado:13/06/2018

DOI: <http://dx.doi.org/10.5557/IIMEI9-N16-049090>

Descripción propuesta: Moro Cabero, Manuela, 2018. Identificación, caracterización y selección de formatos para la preservación del recurso digital. *Métodos de Información*, 9(16), 49-90.

## **1. Introducción**

Los formatos evolucionan a medida que los desarrolladores identifican e incorporan nuevas funcionalidades para las aplicaciones. Se encuentran supeditados a elevada proliferación y obsolescencia, generando obligaciones para el profesional responsable de la preservación de recursos digitales. Este hecho, obliga a los preservadores a conocer todo tipo de iniciativas y aplicaciones para la identificación y evaluación de su viabilidad. El preservador ha de asumir su conocimiento desde un enfoque de pasado (posible formato empleado), de presente (actividad y viabilidad del mismo) y de futuro (necesidad de mudanza), desde una posición activa, racional y sostenible. Por ende, para afrontar esta polimorfa perspectiva nos preguntamos ¿Qué debe conocer y cómo debe actuar el gestor de colecciones bibliotecarias y de fondos de archivo ante esta situación?

El conocimiento de la estructura del objeto digital, así como el de las aplicaciones e iniciativas existentes para facilitar la caracterización y selección del formato, con anterioridad o durante el ciclo de vida del recurso digital, permitirá al profesional afrontar con mayor desenvolvimiento y garantías su preservación.

En este sentido, el objetivo principal de este estudio, una vez mostrada la necesidad de identificar, describir, valorar, seleccionar, mantener y preservar la estructura del objeto digital durante el tiempo que sea preciso, es el de informar sobre los instrumentos y aplicaciones disponibles para su logro. Se pretende, además, pautar el establecimiento de un inventario, de una matriz y perfil de formatos. A la par, se busca presentar las principales herramientas y aplicaciones de utilidad para su identificación y control.

El ensayo es de naturaleza descriptiva, analítica y comparada. La consulta de estudios científicos sobre el formato permite la sistematización de sus contenidos para aportar conocimiento sobre la naturaleza, valor y principales riesgos a los que se encuentran sometidos desde el enfoque del preservador. Así mismo, se analizan las herramientas mediante su identificación y comparativa de funcionalidades más relevantes. Se sistematizan aspectos operativos para facilitar la aplicación de los conocimientos, en la creación de perfiles e inventarios, así como ante la selección de formatos, mediante la presentación de factores, criterios e indicadores.

El ensayo compara las funcionalidades de las principales herramientas específicas creadas bien para identificar, bien para gestionar y/o controlar los formatos. A su vez, proporciona un conjunto de instrumentos, de criterios y de pautas, todos ellos orientados a facilitar el trabajo del preservador.

Para tal fin se estructura en tres apartados. A la par que se subraya la variedad y riqueza de formatos, se reflexiona, precisamente, sobre el grado de incertidumbre desencadenado ante su multiplicación, así como sobre su grado de obsolescencia (enunciado 1). Dichos aspectos permiten soslayar la necesidad de identificarlos, describirlos y seleccionarlos (enunciado 2). El control de formatos obliga a conocer y analizar las herramientas empleadas por los preservadores, con el fin de conocer las funcionalidades que aportan (enunciado 3)

## **2. El formato. ¿Posición interesante o de incertidumbre?**

La multiplicidad de objetos digitales representa un número incalculable de formatos que el profesional de la información y documentación ha de identificar, analizar o mudar para realizar las correctas recepciones, tanto como para afrontar futuras estrategias de mantenimiento y conservación a largo plazo. En este sentido, Cruz-Mundet y Díez Carrera (2015, p.19) especifican una taxonomía heterogénea de recursos digitales, tanto para imágenes, texto, vídeo, audio, datos geoespaciales, Web, publicaciones multimedia, etc.

El contenido informativo de un recurso digital adopta una u otra estructura, atendiendo al programa y sistema operativo donde ha sido creado, almacenado o visualizado, incluido el modo en que ha sido interpretado, tanto

por un navegador, por ejemplo, como por otra aplicación. Esta cualidad de arbitrariedad es señalada por el National Archives of United Kingdom (2011, p.6) al considerar al formato de archivo como un “método arbitrario de almacenar contenidos digitales en un archivo, lo que permite su posterior recuperación o intercambio con otras personas y equipos”. A su vez, y destacando dicha particularidad, la Librmediante su Programa Nacional de Infraestructura de Información y Preservación (2018a), define los formatos como: “paquetes de información que pueden almacenarse como archivos de datos o enviarse a través de la red como flujos de datos (también conocidos como flujos de bits, flujos de bytes”).

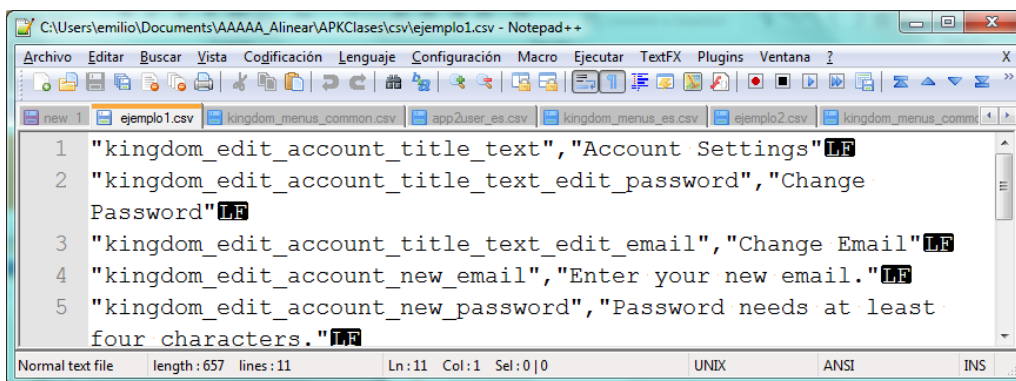
La abundancia de formatos es quizá en el presente, uno de los desafíos mayores al que se enfrentan los gestores de documentos, bien para controlar la creación del recurso, sea para alcanzar la viabilidad de su interoperabilidad, portabilidad, copiado y compatibilidad retrospectiva ante la captura y mantenimiento de ficheros de datos. De igual modo, esta necesidad de control afecta a los preservadores, ante el trazado de un plan preservador y la aplicación de las estrategias formuladas en él. El preservador debe activar el filtrado de conversión de formatos o, en su caso, de migración de la información del recurso, de su replicado, de la emulación del entorno donde se produce o reproduce.

En entornos electrónicos supeditados a constantes cambios, la obsolescencia de formatos, dificulta su reconocimiento, complicando el trabajo diario de gestores de documentos y de conservadores de recursos informativos digitales. Por lo tanto, el reto de controlar tanto el incremento de formatos como su obsolescencia contribuye a ahondar en la preocupación presentada por los profesionales de la información, dado que estos deben afrontar en su trabajo la identificación, valoración y selección de los formatos, en cuanto que conforman la estructura del objeto digital. Esta obsolescencia, no siempre es observada como un desafío por algunos especialistas (Rosenthal 2015), si bien, debe de ser reconocida, en su caso.

Entre los elementos integrantes del recurso digital se reconoce el objeto de datos configurado en un formato y dotado de determinadas propiedades y características. Los datos son interpretados mediante programas de software. Estas aplicaciones logran con éxito la lectura e interpretación de los datos

basándose en su formato, por ende, el formato ocupa una posición interesante entre los datos (cadena de caracteres) y el contenido digital. De facto, el nexo que materializa un contenido digital es el formato. Este viene a representar un rol de intermediario entre el conjunto de “ceros y unos” (código binario) y el manifiesto recurso digital. En atención al tipo de recurso, aún con el mismo contenido informativo y heterogénea u homogénea apariencia, el formato difiere (o puede diferir). La razón radica en la posibilidad de disposición de ficheros u objetos de datos con una estructura predominante: ceros y unos. Si bien, la interpretación de los mismos, “a priori”, es posible efectuarla en situaciones diferentes y con resultados distintos por lo que atendiendo al lenguaje adoptado será factible el empleo de unos u otros formatos y aquello que, en principio, contemplamos como una cadena de caracteres -01001110- llegue a ser visualizado y expresado de modo diferente.

Por ejemplo, un fichero con idéntica secuencia de ceros y unos (idéntica información) al ser interpretado por un navegador como Firefox o por Notepad++ se presenta de modo bien distinto. En este caso, los programas de edición interpretan y manipulan los datos, mostrando la información de diferente modo. Así, en la figura 1 se muestra un fichero en formato “CSV”. Este es un formato abierto que expresa valores separados por comas, el cual fue ideado para intercambiar datos entre gestores de hojas de cálculo, tales como. Excel, Calc, etc. Este es interpretado por dos programas. Como es observable en la figura, el fichero es abierto por el programa Notepad++- editor de texto plano, la coma muestra el cambio a otra columna. No obstante, el texto se entrecomilla. El carácter fin de línea -LF- indica una nueva fila. Se aprecian tres saltos de línea.



```
C:\Users\emilio\Documents\AAAAA_Alinear\APKClases\csv\ejemplo1.csv - Notepad++
Archivo  Editar  Buscar  Vista  Codificación  Lenguaje  Configuración  Macro  Ejecutar  TextFX  Plugins  Ventana  ?
new 1  ejemplo1.csv  kingdom_menus_common.csv  app2user_es.csv  kingdom_menus_es.csv  ejemplo2.csv  kingdom_menus_comme
1  "kingdom_edit_account_title_text", "Account Settings" LF
2  "kingdom_edit_account_title_text_edit_password", "Change
Password" LF
3  "kingdom_edit_account_title_text_edit_email", "Change Email" LF
4  "kingdom_edit_account_new_email", "Enter your new email." LF
5  "kingdom_edit_account_new_password", "Password needs at least
four characters." LF
Normal text file  length: 657  lines: 11  Ln: 11  Col: 1  Sel: 0 | 0  UNIX  ANSI  INS
```

Figura 1- Fichero de texto editado mediante editor NotePad++

Fuente: elaboración propia.

Por el contrario, en la figura 2 se observa idéntica información aunque visualizada con otro programa. Así, la línea uno observada en la figura 1 con el entrecomillado “*Kingdom...*”, la localizamos en la figura 2, en la columna uno de la fila uno. De igual modo, el texto “*Account Settings*” de la figura 1, lo observamos en la figura 2, en la misma línea, aunque en la columna “B” de la misma fila. Dado que en la figura 1, se observa seguidamente un salto de línea –LF, la información se somete a un salto de línea, pasando en la figura 2 a otra fila.

	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		

Figura 2- Fichero de texto editado mediante aplicación de hoja de cálculo CALC.

Fuente: elaboración propia.

Las figuras nos informan sobre el impacto de un formato, sea observado con un programa de texto plano (figura 1) o mediante un programa de hoja de cálculo (figura 2). Esta representación es claramente indicativa de la necesidad de control de la estructura del objeto digital.

Por otro lado, la multiplicación de formatos implica la existencia de *miles de extensiones* de ficheros. La extensión de un fichero nos informa sobre qué tipo de datos y de archivo se hallan en funcionamiento y qué programa asociado abre dicho archivo. La diversidad de recursos con posibilidades de formatos innumerables, conocidos o no identificados, torna complejo el control de los formatos, siendo, a su vez, dicho control esencial. Por ende, el reto es doble: variedad de formatos (se computan en miles) y obsolescencia (favorece el desconocimiento de los formatos).

La categorización de los formatos facilita su reconocimiento. De facto, si disponemos del objeto de datos y de la información de representación de ese objeto, podremos controlar su obsolescencia y gestionarla. Esto es, reconocer su estructura.

A su vez, el control de los formatos se torna más complejo al considerar que estos inciden en la apariencia del contenido, motivado por la evolución de aplicaciones y entornos tecnológicos. Esto es, en ellos incide la variabilidad que suponen las versiones y los sistemas operativos o/y aplicaciones con los que se opere, aspecto que contribuye a multiplicar el impacto de proliferación y obsolescencia (en cuanto a versiones existentes).

En suma, se constata que los formatos expresan el paso del tiempo intergeneracional de entornos electrónicos. En calidad de preservadores, debemos identificarlos, conocer sus propiedades, sus riesgos, su viabilidad, atendiendo a los usos y necesidades y considerando, además, su potencial en la conversión, migración u otras estrategias preservadoras.

### **3. El control de formatos en la organización.**

En este enunciado se fundamenta la necesidad de control de formatos (2.1) mediante su identificación, caracterización y valoración para la selección. Dicho control es necesario efectuarlo en el desarrollo de un plan de preservación para los recursos digitales. A tal fin, se sustenta dicho control en especificaciones de alcance internacional y en estudios metodológicos de expertos en preservación. En un segundo epígrafe se profundiza en la identificación de los formatos, detallando los elementos e instrumentos para ejecutar dicha tarea (2.2). Caracterizados los formatos, se ahonda en el siguiente subepígrafe (2.3) en su valoración, mediante la enumeración y descripción de criterios empleados para tal fin. Finalmente, se establece un último subepígrafe (2.4) en el que se determinan pautas y experiencias sobre la selección de los formatos, considerando esta actividad, vinculada a los resultados obtenidos previamente ante su identificación y valoración.

#### **3.1 Fundamentación**

El objeto digital debe de ser codificado y decodificado para la obtención de información de contenido. El primer modo de aproximarnos a él es mediante su identificación. Por ende, en toda organización debe realizarse un catálogo de recursos digitales, donde se expresen, atendiendo a categorías u otros criterios adoptados, volumen, localización, ubicación en el entorno tecnológico de producción, mantenimiento o visualización, necesidades y usos del recurso, naturaleza del recurso (esencial), constatación de datos personales, grado de confidencialidad de la información contenida, permisos requeridos, valoración y disposición asignada, etc., además del formato o formatos en el/los que se presenta. En dicho catálogo deben quedar explicitadas las respuestas a interrogantes sobre el recurso digital, tales como: ¿Quién lo crea/transforma o usa?; ¿Dónde se crea, transforma y usa?; ¿Cómo se usa y usará?; ¿En qué se crea, transforma, visualiza y utiliza?; ¿Durante cuánto tiempo?; ¿Cuál es su tipología o/y categoría?; ¿Dónde se ubica?; ¿Qué estructura presenta?; ¿Durante cuánto tiempo será válida?, etc.

La estructura del recurso digital ha de ser siempre investigada, dado que el formato es parte identitaria. Mediante el conocimiento de la estructura del fichero se identifica el formato, disponiendo de mayores opciones para su control y conservación. Dicho conocimiento debe ser considerado un requisito relevante. Como tal es observado en todo plan de preservación. A modo de ejemplo referenciamos entre los procesos y procedimientos que el Subcomité 11 del TC46 de ISO (ISO/TC46/SC11/WG7 Digital Records Preservation; 2010, p.9), establece en su abreviada guía para elaborar un plan de preservación, el siguiente, entre otros: "la identificación de los formatos y soportes de almacenamiento", igualmente, en todo plan de preservación se deben de incluir las limitaciones técnicas que son referenciadas en dicha guía (p. 7) siendo enumeradas la: "obsolescencia de soportes, de formatos de fichero, de software y de hardware". Precisamente, estas limitaciones técnicas son calificadas en la directriz como desafíos.

Esta condición se observa de modo específico en cualquier proyecto metodológico de aplicación de una estrategia preservadora. Así, en las pautas metodológicas de la norma UNE-ISO 13008 (2013) se incluye la necesidad de identificar fuentes de producción (incluidos los formatos) de los recursos digitales sujetos a conversión, aunque también, de seleccionar los formatos de destino considerando, estructura de los datos, funcionalidades disponibles,



necesidades de uso del recurso digital y recursos. En la fase segunda establecida para un proyecto de migración/conversión, en la que se ejecuta el proceso propiamente dicho, se incluye la actividad definitoria de los recursos, incluida la específica a su formato.

Así mismo, J. Voutssás (2013, p.148), en sus diez recomendaciones para la preservación del patrimonio digital personal, enuncia la necesaria selección de formatos para maximizar la accesibilidad. A modo de mandatos de ley, establece en primer lugar la catalogación de objetos digitales, desde un enfoque amplio, incluido el tecnológico y, en segundo término, la selección de formatos con el objetivo de optimizar la finalidad preservadora, esto es, de disponer la información accesible, usable y reutilizable en el tiempo.

De este modo se sustenta la necesidad de identificar los recursos digitales, sus formatos, el entorno tecnológico con el que fueron creados, incrementados o modificados, así como su grado de obsolescencia, las versiones, los riesgos a los que se ven sometidos, las propiedades significativas de los mismos, etc.

A su vez, considerando los postulados regulados en UNE-ISO 30301 (2011, p.21) en el que se nos señala la necesidad de actuar sobre el recurso digital desde la fase de creación, para un recurso de archivo, el control del formato queda reflejado en el objetivo A1.3 (decidir en qué forma y estructura se deben de crear y capturar los documentos) y en su control que seguidamente se expresa: identificar forma y estructura. Además, se enfatizan otros objetivos para las tecnologías a emplear (en la fase de creación) y las aplicaciones informáticas implementadas (en la fase de control) para facilitar la documentación y el control tecnológico.

La identificación del formato se efectúa mediante el inventariado de formatos, expresándose en un *inventario de formatos*, el cual ha de vincularse estrechamente al catálogo de objetos o recursos digitales. Identificados los formatos existentes en una organización, será preciso establecer una selección de los mismos, atendiendo a criterios de valoración o a las limitaciones impuestas por repositorios y administraciones (ejemplo, sobre portabilidad de ficheros e interoperabilidad). A su vez, si los recursos digitales precisan de un mantenimiento y conservación a medio y largo plazo, será necesario controlar la viabilidad de sus formatos, expresándose el nivel de preferencia en cuanto uso mediante el *perfil del formato*.

De este modo, el profesional gestor de los recursos digitales, ante acciones de identificación, selección y control de formatos, tal y como se refleja en la tabla, deberá crear determinados instrumentos y disponer de aplicaciones auxiliares cuya adopción facilitará su trabajo en determinadas tareas relacionadas con el control de los formatos.

<b>Acciones del gestor de recursos digitales</b>	<b>Instrumentos y aplicaciones</b>
Identificación de los recursos digitales existentes	Catálogo de recursos digitales
Identificación de formatos existentes en la organización.	<i>Inventario de formato</i>
Conocimiento de propiedades del formato Valoración y selección de formatos  Control del riesgo vinculado a viabilidad de formatos.	<i>Matriz del formato</i>  Factores; Criterios e indicadores ante/para la selección  <i>Perfil de formatos</i>
Control del uso e informaciones diversas sobre los formatos (ante recursos que deben de ser capturados, migrados, convertidos, importados, exportados, emulados, digitalizados, copiados, etc.).	<i>Aplicaciones</i> varias con amplias funcionalidades de control: DROID, JHOVE2, FIFT, METADATA, XENA, PRONOM...; <i>Funciones</i> de planificación, de almacenamiento, de consulta en el repositorio. <i>Normas</i> de formato

Tabla 1- Sistematización del trabajo del profesional en lo referente a formatos.

Fuente: elaboración propia.

Se computan como veremos múltiples aplicaciones informáticas dotadas de heterogéneas funcionalidades que contribuyen al control del recurso digital, incluida su estructura, atendiendo a situaciones dispares en su ciclo de vida, de utilidad para su identificación, para el control del mismo en el momento de

importaciones o exportaciones, ante captura en un repositorio, ante la aplicación de una estrategia de preservación (copiado, migración, conversión, emulación, digitalización, entre otras) o bien, ante su consulta o reproducción (requiriendo un formato en atención a demanda, controlado por funciones del repositorio).

Desde esta perspectiva, algunas de estas aplicaciones han de contemplarse como fuentes secundarias de información sobre el formato, siendo de gran utilidad para gestores de colecciones y fondos y muy especialmente para los preservadores.

### **3.2 Identificación de formatos**

La identificación del formato ha de realizarse con suficiente detalle descriptivo precisando disponer de la máxima información sobre el mismo. En un catálogo de recursos digitales, por regla general, se incluyen datos informativos sobre nombre extenso del formato y a ser posible su extensión. Ambos aspectos aunque importantes son insuficientes para establecer valoraciones del formato. De ahí la importancia de disponer de una herramienta en la que se describa cada formato en extenso, la cual se conoce como inventario de formatos y es resultado de la actividad de inventariado.

Un inventario de formatos puede crearse mediante una sencilla base de datos, donde se registre información sobre los aspectos descriptivos del mismo que faciliten su identificación además de servir de apoyo para otras acciones caracterizadoras y evaluativas del mismo. Así, entre los elementos descriptivos se enumeran los siguientes: número unívoco asignado a cada formato, nombre extenso del formato, extensión, versión o versiones disponibles, naturaleza o tipo del formato (abierto o propietario), caracterización del desarrollador y, en su caso, del mantenedor del formato; caracterización sobre su versatilidad y utilidad (a través de sus propiedades y estimación de vida); opciones de preservación o viabilidad del mismo (opciones para la conversión, softwares que lo abren, archivan, permiten consulta, interacción, exportación e importación, etc.), principales bibliotecas o centros que lo emplean, Tipos MIME aplicables (información sobre el contenido de un Archivo: audio, aplicación, mensaje, modelo, texto, multi-parte), enlaces clave relacionados

con el formato (donde se aporta información complementaria), riesgos o factores que inciden en él y tipo de mitigación del riesgo observada, información sobre si está sujeto o no a licencias y sobre su problemática, así como sobre sus propiedades más significativas: calidad, coste, funcionalidad, resiliencia, flexibilidad, sostenibilidad, etc.

La práctica de inventariar es apreciable en la E-Administración española, donde la norma técnica de interoperabilidad Catálogo de estándares, (España. Ministerio de Hacienda y Administración Pública; 2017, pp 452 y ss.) muestra un listado de estándares mínimos necesarios para la interoperabilidad en los que se señalan los elementos siguientes: eslabón de la cadena de interoperabilidad con el que se relaciona, categoría funcional en la que se enmarca el estándar, nombre formal, tipo, versión, extensión, tipo y estado, de admitido o no. Aunque al tratar de inventariar estándares, no presenta un inventario al uso, sí que nos orienta sobre formatos en uso o desuso para facilitar la interoperabilidad de los ficheros entre administraciones.

Por otro lado, para cada formato incluido en el inventario de formatos sería altamente recomendable aportar un vínculo con el número unívoco del (los) recurso(s) digital(es) a(l) los que afecta (atendiendo a las versiones, por supuesto). Esta vinculación debería ser extensiva también para el catálogo o inventario realizado sobre el entorno tecnológico de creación, manipulación y almacenamiento de los activos digitales. Por tanto, a nuestro entender resulta imprescindible relacionar formato, objeto digital y entorno donde se crea, usa, mantiene y preserva con sus instrumentos de control, tal y como se expresa en la figura siguiente (figura 3):

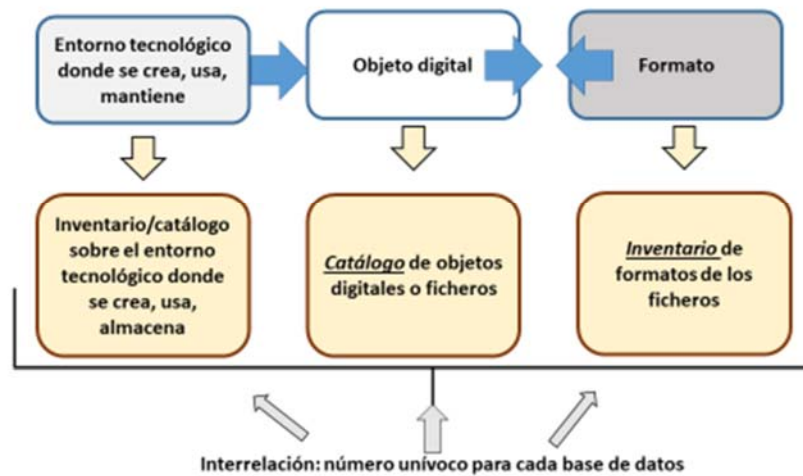


Figura 3. Relaciones establecidas del formato, recurso digital y entorno para su control. Fuente: elaboración propia.

La información expresada en el inventario para cada formato permite su identificación y descripción, auxilia en su valoración y facilita la selección de aquellos más adecuados atendiendo a una serie de criterios e indicadores, los cuales pasamos a analizar.

### 3.3 Valoración de formatos

El proceso de valoración de los formatos es prioritaria por múltiples razones, de las que se destacan cuatro. La primera, es que el incremento de formatos en una organización implica mayores esfuerzos económicos y de control de los objetos digitales, de ahí, la necesidad de normalizarlos y reorientarlos bajo la máxima “cuantos menos, mejor”. La segunda razón se relaciona con el grado de obsolescencia de entornos y objetos digitales, el cual incide en el control de los formatos incrementando, igualmente, esfuerzos, tiempos, recursos económicos y funcionalidades en las aplicaciones y unidades preservadoras. La tercera motivación se justifica sobre la existencia de formatos (propietarios o semipropietarios) que están sujetos a licencias, obligaciones que dificultan la sustentabilidad de las unidades preservadoras, habida cuenta de que éstas, en principio suelen disponer de recursos muy limitados para asegurar la disponibilidad de la información en el tiempo. La cuarta motivación, se vincula a la funcionalidad y calidad del formato para operar y responder a los

requisitos de negocio, así como a las necesidades informativas de futuras comunidades específicas de usuarios.

En principio, cualquier visualización de ficheros en la pantalla viene a parecerse un listado de “líneas similares” (a priori, todos se asemejan). El reconocimiento de su formato nos permitirá editarlos, visualizarlos, importarlos, exportarlos, convertirlos, etc. Además, sería aconsejable, igualmente, disponer de información sobre su coste, su calidad, su sostenibilidad, etc. ante la decisión sobre cuál es el mejor para mantener o para conservar la información (sea en un ciclo de vida corto, medio o a largo plazo).

La selección de formatos nos permitirá actuar con juicio y criterios científicos. Seleccionar implica reducir riesgos ante obsolescencia, ante abundancia de formatos, ante funcionalidad, ante necesidad de la disponibilidad de la información del recurso. Por ende, la selección de formatos se relaciona estrechamente con su control y viabilidad, no sólo en el tiempo, sino que también desde un enfoque de rentabilidad al considerar costes y sostenibilidad.

En la figura siguiente (figura 4-Relación de criterios para la selección de formatos) se reúnen destacados criterios aconsejados por diversos especialistas ante la selección del formato:



Figura 4- Relación de criterios para la selección de formatos

Fuente: Elaboración propia

Como es observable en la figura, han sido consignados ocho criterios que seguidamente son comentados:

El primer criterio se vincula a la utilidad informativa del recurso digital. El conocimiento de las necesidades informativas presentes y futuras de las comunidades específicas, facilitará determinar su destino o uso. Constatamos que puede diferir la necesidad de uso del recurso para una visualización (cuyo nivel de reproducción medio o bajo sería suficiente) de aquella que precisa su conservación a medio o largo plazo y con fines de explotación o reutilización. Ante todo, debemos de ser conscientes de que audiencias heterogéneas precisan ficheros para acciones divergentes. En este sentido, el National Archives of United Kingdom (2011, p.9), señalan los siguientes usos: crear la información, encontrarla, abrir un fichero, trabajar con la información contenida, comprenderla y verificarla. Así pues, los usos a realizar en la plataforma pueden ser diversos, desde creación, edición, provisión de lectura, hasta de búsqueda o revisión.

El segundo aspecto reseñado alude a la necesidad de valorar el grado de portabilidad de los ficheros (exportación e importación) así como de interoperabilidad entre administraciones diferentes. Este hecho es relevante, dado que las organizaciones se interrelacionan con la sociedad en la que se hallan imbricadas, actúan como sistemas abiertos a cuyo reflejo no son inmunes los ficheros de datos y los flujos de información.

Un tercer criterio a considerar es el grado de propiedad del formato. Esto es, si se trata de un formato abierto, semiabierto o propietario. Dicho grado puede determinar la necesidad de conversión, igualmente, atendiendo a las necesidades de uso del objeto digital. En este sentido, los Archivos del Estado australiano de Queensland (2018) especifican riesgos, en atención al grado de propiedad. Así, para un formato abierto el riesgo del formato es bastante bajo, frente a un formato cerrado o semi-cerrado, dado que la propiedad intelectual del mismo está públicamente disponible o responde en numerosos casos a estándares nacionales e internacionales. La situación de formato semi propietario, implica que está sujeto a derechos compartidos aunque su

propiedad es de naturaleza privada. Se deja patente cómo los formatos no propietarios responden bien a la sostenibilidad de los datos a largo plazo y aportan mayores posibilidades de resistencia y de compatibilidad a los cambios de futuro, frente a las restantes situaciones de propiedad.

Un cuarto factor relevante es la constatación de la existencia de un amplio grado de conocimiento sobre el formato. Esto es, la existencia de un inventario de formatos o en su caso, la constatación de obtener información sobre el mismo en aplicaciones específicas o estándares reguladores total o parcialmente.

Un quinto criterio estrechamente relacionado con el anterior es la disponibilidad de fuentes informativas sobre el formato o/y estructura del recurso en cuanto a normas, documentación del desarrollador o mantenedor, documentación vinculada al entorno tecnológico donde se crea, visualiza, etc. No siempre se encuentra disponible documentación. Sin la existencia de un inventario de formato o, en su caso, de documentación sobre el mismo procedente de los desarrolladores o mantenedores, la opción para su reconstrucción pasa por el conocimiento de aplicaciones específicas que serán comentadas en el epígrafe siguiente.

Un sexto criterio destacable es el conocimiento sobre el grado de pérdida del formato; el cual resulta esencial desde una perspectiva de sostenibilidad. Disponemos de formatos que se emplean para reproducción con baja resolución; un almacenamiento con baja resolución imposibilita el retorno hacia las propiedades anteriores del formato. Por ejemplo, un fichero de audio, al reproducirlo en MP3 se produce una pérdida de datos no siendo posible un retorno a la situación de partida. Acontece, igualmente, con los formatos JPEG. Por el contrario, un formato TIFF no conlleva pérdida.

A menudo, en nuestra área se habla de compresión, pero los ficheros comprimidos, en principio no actúan como aquellos de representación donde suele haber una pérdida de datos sin retorno. Un fichero comprimido, *a priori*, tiene opción de retorno sin pérdida alguna de datos. Este hecho es verificable ante el desempaquetado de un fichero comprimido (del que lógicamente previamente se ha conservado una copia máster u original) y su comparación con el original. Se observa un retorno del 100%. Suele calificarse con el



término inapropiado de “comprimir” a la operación referida a lograr espacio de almacenamiento y rapidez de recuperación, puesto que los formatos de representación (con pérdida) aportan esas ventajas. Por ejemplo, para disponer de mayor capacidad de almacenamiento en nuestro móvil, solemos reducir la resolución de las fotografías. Reducir resolución implica pérdida de calidad en la imagen y ganancia en capacidad de almacenamiento.

Dada su importancia, ante la selección de formatos se recomienda seguir las siguientes reglas de oro:

- Seleccionar formatos sin pérdida, siempre, para creación y almacenamiento: afectan al original y a las copias de seguridad almacenamiento.
- Seleccionar formatos con pérdida (de representación y resolución) únicamente para usos de entrega/acceso temporal.
- Nunca emplear formatos con pérdida en los ficheros para almacenamiento a largo plazo.

Un séptimo criterio a considerar es si el formato admite (o no) en el encabezamiento del fichero metadatos. Esto resulta importante para la gestión de metadatos. Igualmente, resulta interesante comprobar si en las conversiones de estos metadatos son exportados sin pérdida de información. Los metadatos técnicos y descriptivos del fichero se muestran esenciales. Existen aplicaciones que facilitan este control, tal y como se verá en posterior enunciado.

Finalmente, el octavo criterio destacado se vincula al análisis de las propiedades significativas del recurso y del formato. Especialmente, en cuanto a contenido del recurso digital (ver qué tipo de alteraciones se producen ante una exportación, etc.) y su entorno tecnológico. Al respecto, cabe señalar que los formatos propietarios tienen mayor dependencia de los entornos tecnológicos. Por ejemplo, es posible encontrar formatos que aunque en principio son abiertos para su visualización, precisan de software propietario. Algunas instituciones han trabajado someramente estas propiedades, disponiéndose de información sobre ellas para categorías de recursos digitales. Tal es el caso de la Biblioteca del Congreso angloamericano o del Archivo Nacional de Reino Unido.

Así, en el Archivo Nacional del Reino Unido, Adrian Brawn (2008) desarrolló una guía de criterios para la selección de formatos en la que se señalan los siguientes: ubicuidad, tolerancia, divulgación de la información, calidad, estabilidad, identificación y validación sencilla, derechos de propiedad intelectual, consistencia de los metadatos, complejidad, interoperabilidad, viabilidad y re-usabilidad. La guía se encuentra accesible en línea, tal y como es observable en la referencia.

Por otro lado, la Biblioteca del Congreso Estadounidense (Library of Congress, 2018), mantiene una herramienta mediante la que es posible consultar propiedades significativas de los formatos, atendiendo a las exigencias específicas para categorías documentales. Las características señaladas para cada categoría, aportan información, tanto para creadores y desarrolladores de los recursos y del software, como para los usuarios, presentes y futuros. A modo de ejemplo, referenciamos características de calidad detalladas para imagen fija (Library of Congress, 2018b). Así mismo, señalamos aquellas características exigibles para preservar el contenido de los recursos digitales textuales: además de la comprensión del contexto de creación del recurso y de los cambios de su entorno tecnológico que puedan afectar a la estructura y a la presentación, se deberían de constatar la capacidad para evaluar el valor del documento como evidencia histórica, la capacidad para recuperar y analizar segmentos independientemente de la representación original, capacidad para navegar a ítems referenciados y entender aspectos cronológicos de las relaciones, capacidad de reutilizar el contenido. Asimismo, se subraya la necesidad de control de los cambios de su estructura (esto es, de su formato). Para los recursos de imágenes se especifican características de calidad.

De igual modo se especifican factores de sostenibilidad tales como: divulgación del formato, grado de adopción, transparencia del mismo, auto-detección, dependencia externas, impacto de las patentes y mecanismos de protección técnica (Library of Congress, 2018). Algunos de ellos, han sido ya comentados. No obstante, merece la pena destacar de entre los factores, aquellos no mencionados, tales como la cualidad de transparencia, refiriéndose al grado en que la representación digital facilita el análisis directo o mediante herramientas disponibles de tipo básico y mediante el empleo de un solo

editor. De igual modo, se ha ahondado en la auto-documentación al enunciar la vinculación con los metadatos y sí resulta interesante destacar el factor sobre mecanismos de protección técnica, recordando que opciones de encriptado o de incorporación de dispositivos protectores de la propiedad intelectual (mediante la restricción del uso del recurso) dificultan la preservación a largo plazo de los objetos digitales.

De la enumeración de criterios realizada se desprende que la valoración para la selección de un formato no resulta una tarea sencilla. Más bien, exige la recogida de información en extenso sobre dicho formato, además de sobre otros elementos como puede ser el conocimiento del ciclo de vida del recurso digital, las necesidades de uso del mismo, el entorno tecnológico, etc. Sin suficiente conocimiento sobre la estructura del recurso digital resulta difícil adoptar una decisión reflexionada. La decisión adoptada ha de estimar, igualmente, otros indicadores vinculados al riesgo al que se encuentra sometido un formato, y siempre desde una finalidad reductora del mismo.

La disminución de los riesgos a los que se encuentra sometido el formato no es baladí. Toda mitigación, por ende, comienza considerando los aspectos y factores anteriormente citados. Destacamos, a modo de sistematización:

- El grado de uso del formato, tanto a nivel internacional como a nivel nacional o local. Cuanto más generalizado sea el uso del formato, mayores posibilidades de pervivencia del mismo. No obstante, formatos usados a nivel mundial como fue el utilizado por el programa *Word Perfect* han desaparecido rápidamente, hasta de nuestra memoria.
- Se deben considerar sus dependencias sobre el hardware y software existente en la organización y el nivel de interoperabilidad; Conservar un formato muy dependiente a corto plazo puede ser rentable, aunque no siempre lo es a largo plazo.
- La apertura y la normalización pública de las especificaciones de formato y el impacto de la patente; Por ejemplo, formatos bajo norma ISO, como WARC (ISO 28500:2017), (para almacenamiento de páginas Web) o PDF aportan mayor seguridad.

- La transparencia del formato, incluyendo compresión y legibilidad humana; Cuanto más transparente, mejor. No siempre es factible este aspecto.
- Conocer qué soporte de metadatos está disponible incluyendo la auto-documentación; Estos soportes no siempre son compatibles con los formatos seleccionados. Al respecto, igualmente, ha de considerarse, si los metadatos incluyen gestión de derechos digitales y el grado de protección técnica y de encriptación; Pudiera darse el caso de que no se paguen derechos de licencias por el software y sí por los metadatos que lleva incrustado el recurso.
- Conocer el grado de compatibilidad del formato hacia adelante y hacia atrás y con otro software o hardware, para evitar pérdida de datos. Al respecto, es conveniente recordar que una disminución de capacidad en el almacenamiento, no siempre significa calidad en la integridad de los datos. También hay que controlar las versiones de un formato sobre sus compatibilidades retrospectivas.
- La robustez del formato, incluyendo su complejidad; A tal fin, lo más oportuno es considerar las recomendaciones de grandes sistemas de bibliotecas y Archivos, sobre todo, en materia de preservación. El perfil que ofrecen de los formatos para preservación a largo plazo puede ser una buena orientación.
- La habilidad para manipular y reutilizar el contenido del fichero; Lógicamente, el contenido del fichero es lo esencial del activo digital. El ejemplo dado anteriormente sobre MP3 es claro en cuanto a la disminución de la calidad en la reutilización. Un fichero TIFF por ejemplo no tiene pérdida de datos y posibilita la reutilización.

De facto, son ya numerosos los centros que han creado sus propios criterios y determinado factores para facilitar la valoración. Así, el National Archives of United Kingdom (2011, p.10) establece para la evaluación en un contexto de continuidad digital, los factores siguientes: capacidad para el cumplimiento de los requisitos de negocio, calidad en el almacenamiento de la información, resiliencia de la información a largo plazo y flexibilidad para adaptarse a requisitos cambiantes. En su detalle descriptivo ha formulado los siguientes criterios:

<b>Relación de factores</b>	<b>de</b>	<b>Criterios</b>
Capacidad		Obligada: <i>disponibilidad en diferentes sistemas, búsquedas, transformación a Pdf, apoyo de gráficos, etc.</i> Deseable.
Calidad		Precisión. Pérdida.
Resiliencia		Ubicuidad. Estabilidad. Recuperación.
Flexibilidad		Interoperabilidad. Aplicabilidad.

Tabla 2. Relación de criterios ante la continuidad digital

Fuente: National Archives of United Kingdom (2011, p.10). Elaboración propia

En esta línea, la Harvard Library digital Preservation (2017) dispone de una matriz mediante la cual es factible valorar aspectos tales como sostenibilidad, costes, funcionalidad y calidad, además de otros elementos, a partir del establecimiento de una escala donde se identifican criterios e indicadores para cada uno de los citados factores. A su vez, los resultados son reflejados mediante gradaciones de valor (pobre, moderado, bueno; bajo, medio, alto) y uso de distintos colores (verde, amarillo o rojo). Seguidamente, se expresan en la tabla para cada factor los criterios que han sido acordados en esta entidad.

<b>Relación de factores</b>	<b>de</b>	<b>Criterios</b>
Sostenibilidad del formato		Adopción. Divulgación (conocimiento del formato). Transparencia. Auto-documentación. Dependencias externas. Impacto sobre patentes. Soporte para mecanismos de protección.
Costes		Mantenimiento del entorno para el acceso. Requisitos para la descompresión

Relación de factores	Criterios
	/almacenamiento raw. Requisitos de almacenamiento con máxima reducción.
Funcionalidad y calidad	Apoyo cadena de color. Máximo número de bytes por cadena de color. Resolución Máxima. Capacidad para codificar en modo compresión (con/sin pérdida). Ratio de compresión (reducción). Sub-muestreo de croma máxima. Máxima profundidad de bytes en general.
Otros:	Empaquetamiento. Apariencia de la imagen.

Tabla 3- Relación de criterios para cada factor

Fuente: *Harvard Library Digital Preservation* (2017) Elaboración propia.

En la tabla siguiente se reflejan los indicadores establecidos para algunos de los criterios integrantes del factor “sostenibilidad del formato”, a modo de ejemplo sobre la profundidad alcanzada por esta matriz (véase tabla 4).

Relación de factores	Criterios	Indicadores
Sostenibilidad del formato	Divulgación (conocimiento)	<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>
	Impacto de patentes	<ul style="list-style-type: none"> <li>• Restricciones legales que afectan al nuevo uso o a largo plazo.</li> </ul>
	Auto-documentación	<ul style="list-style-type: none"> <li>• Tolerancia de errores.</li> <li>• Metadatos técnicas.</li> </ul>

		<ul style="list-style-type: none"> <li>• Metadatos descriptivos.</li> </ul>
	Transparencia	<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>

Tabla 4. Relación de indicadores de la Matriz

Fuente: *Harvard Library Digital Preservation*. (2017) Elaboración propia.

En dicha tabla no han sido incorporados todos los criterios, no obstante, se señalan para algunos de ello, y más concretamente, para los factores de sostenibilidad, indicadores específicos para los criterios de divulgación, impacto de patentes, auto-documentación y transparencia. El objetivo es mostrar el grado de profundidad que proporciona esta matriz en la que la escala para la evaluación de formatos es muy amplia e incluye en última de sus columnas caracterización detallada mediante el empleo del color para cada indicador.

Así, si tomamos como referencia para el factor sostenibilidad y criterio divulgación, el indicador formulado como “grado en que una a especificación sobre el formato completo y comprensible”, se señalan los parámetros (gradientes de color) como sigue:

- Verde: disponible mediante fuente auténtica.
- Amarillo: disponible pero procedente de una fuente anónima o no propia.
- Rojo: no disponible.

A su vez, se indica el método para valorar: “basado en la provisión/o no de una copia de la especificación del formato” (Formulario de matriz de formato).

El reconocimiento y uso de factores y de criterios, así como la formulación de indicadores fundamentan la selección de formatos.

### **3.4 Selección de formatos**

La selección del formato se ha de efectuar una vez ejecutada una valoración racional, mediante el empleo de una matriz dotada de una sólida escala. Esta matriz, ha de ser contemplada como el instrumento básico que facilitará el perfilado de los formatos para una organización.

Generalmente, ante el desarrollo de un *perfil de formatos* se proponen dos categorías de formatos: formatos preferidos y formatos aceptados. El perfil de formatos describe tanto la relación de formatos preferidos como de aquellos aceptados y en él se incorpora información sobre cada uno de los clasificados en una u otra modalidad. Se incluyen elementos informativos extensibles incluso a riesgos y mitigación, licencias, etc. Es posible consultar, igualmente, el perfil de formatos de organismos que reconocen una tercera tipología de formato: formatos recomendados para transferir o acceder al recurso digital (más a corto plazo). Informan al respecto, por ejemplo, el National Archives and Records Administration (NARA) (2018), la Library of Congress (2018c) de Estados Unidos<sup>1</sup> los Archivos Nacionales de Australia (National Archives of Australia, 2018)<sup>2</sup>, etc.

En la figura (figura 5) se representan las categorías establecidas para un perfil, destacando los niveles de preferencia del formato, bien sea preferido o aceptado, bien para transferencia de ficheros, en algunos de los casos. La relación suele establecerse por categorías de recursos digitales (textual, imagen, audio, Web, etc.) y en ella, para cada formato se incorporan elementos descriptivos variables atendiendo a cada propuesta. No obstante, en la figura se incluyen elementos recomendados en el sitio específico sobre perfil de

---

<sup>1</sup> Este centro dispone de una fuente actualizada anualmente realmente interesante sobre los formatos recomendados, en atención a la tipología de los recursos, diferenciante entre formatos preferentes y menos preferentes.

<sup>2</sup> En este centro, los formatos de archivo preferentes y aceptables para la preservación propuestos se basan en estándares abiertos.



formatos de la Harvard Library Digital Preservation (2018).



Figura 5- Elementos descriptivos de un perfil de formato.

Fuente: National Archives and Records Administration (2018). Elaboración propia.

La decisión entre formato preferido, aceptado o para transferencia o acceso inmediato no resulta sencilla. No obstante, considerando factores, criterios y recomendaciones es posible construir herramientas adaptadas para una organización basada en la formulación de indicadores y en la aplicación de una escala que faciliten esta tarea.

Conocida la necesidad de identificar y controlar los formatos de archivo, así como las herramientas necesarias para describirlos (inventario de formato), valorar sus factores y riesgos (matriz), e informar sobre la elección del formato preferente y aceptable (perfil de formatos), resta analizar aquellas iniciativas caracterizadoras y de utilidad para un control racional de los formatos. Estudio que se realiza en el siguiente enunciado.

#### **4. Herramientas para trabajar la preservación de los formatos.**

En este enunciado se identifican las funcionalidades más importantes para trabajar la preservación de los formatos (4.1) y se da paso al detalle de aquellas

herramientas que son utilizadas para el reconocimiento e identificación preliminar (4.2), así como a otras herramientas que cumplen diversas funciones de identificación, validación, caracterización y normalización (4.3). Finalmente, se señala enlace a algunas aplicaciones elaboradas por organizaciones importantes (4.4).

## **4.1 Introducción a las funcionalidades.**

Para la creación de inventarios, matriz y perfil de formatos se ha hecho referencia a experiencias elaboradas por diferentes unidades de información. No obstante, existen múltiples iniciativas para trabajar con cierta profundidad los formatos de ficheros tanto en la captura, gestión y mantenimiento del recurso digital como para la preservación. Cada una de ellas dispone de diversas funcionalidades. Destacamos las siguientes: identificación, caracterización, validación, normalización de archivos o ficheros:

1. La función de identificar permite conocer la estructura del objeto de datos. Esto es, facilita la identificación del mismo.
2. La función de validar **constata la consistencia entre** el fichero y sus metadatos. Por ejemplo, para exportar, importar, ante la captura, etc.
3. La función de caracterizar permite obtener sobre el formato los datos técnicos sobre las propiedades del objeto. Esto es, información sobre el formato, el desarrollador, el mantenedor, así como sobre licencias, documentación, riesgos, etc. Facilita la construcción del inventario de formato.
4. La función de normalizar informa sobre la posibilidad de conversión/migración a otro formato aceptado. Esto es, sobre la viabilidad del formato y sobre posibles riesgos, incluso sobre opciones de retrospectiva en caso de retorno. Facilita la valoración y selección de formatos.

En definitiva, las herramientas que disponen de alguna o todas las funcionalidades, responden a los interrogantes específicos sobre el formato del tipo:- ¿Qué formato tiene el objeto digital? ¿Es conocido? ¿Este objeto con este formato es tal cual se me indica? ¿Qué versión del formato? ¿Qué propiedades tiene o son más importantes en este formato? ¿Puedo

convertirlo a otro(s) formato(s)? ¿Cómo de operativo quedará? ¿Habrá pérdida de información? ¿Mudará la apariencia del recurso? ¿Puedo extraer metadatos o chequearlos? ¿Se encuentran incrustados en el fichero o es preciso disponer de otro fichero? ¿Se produce portabilidad completa o demediada? ¿Cuál es su obsolescencia? ¿Quién es su creador? ¿Quién su mantenedor? ¿A qué tipo de licencia está supeditado? ¿Cuál es su impacto ante su conservación a largo plazo? ¿Dispone de documentación o avales sobre su utilización suficientes? Etc.

A nuestro pesar, no todas las herramientas disponibles aportan información suficiente para dar respuestas a los anteriores interrogantes. De hecho, algunas de ellas, apenas aportan funcionalidades, no obstante, señalamos suficientes para facilitar el trabajo en los sub-epígrafes 4.2 y 4.3.

## **4.2 Herramientas de identificación y primeras iniciativas: GDFR/UDFR**

En este sub-epígrafe serán estudiadas aquellas herramientas de reconocimiento rápido de un formato, así como las primeras iniciativas relevantes desarrolladas para el control de formatos facilitando su identificación.

El conocimiento rápido del formato normalmente lo hacemos, principalmente, a través de la extensión del fichero (código). Claro ejemplo para un documento de texto en entorno Windows es la extensión “.docx”. Su uso nos permite, en este caso, una rauda identificación del formato. No obstante, el conocimiento de su extensión no siempre permite averiguar la versión del mismo. Dicho reconocimiento es posible investigarlo mediante su número mágico.

En la figura siguiente (figura 6. Representación de dos modos rápidos de identificar el formato), se especifican ambas modalidades.



Figura 6. Representación de dos modos rápidos de identificar el formato.

Fuente: elaboración propia.

En el ejemplo expresado en la figura 6, la extensión “.GIF” no nos informa sobre la versión. Sin embargo, el número mágico identifica la versión. Se trata de un código que se encuentra en las cabeceras de los ficheros e indica el tipo de formato. Normalmente, es un código conformado de caracteres alfanuméricos que identifican un archivo, fundamentalmente de tipo binario. Ejemplo, para las imágenes en formato “.Gif” suelen incluirse GIF87a o GIF89a; Este código sí informa sobre la versión. Otros ejemplos a verificar para el identificador de la extensión .XML, es la cadena que comienza con “<?xml”; Igualmente, el análisis de un formato de categoría del recurso textual funciona de modo similar a un número mágico, aportando indicaciones mediante etiquetas del tipo de documento: “Doctype svg” (para un fichero svg, por ejemplo).

En algunos casos, la identificación del fichero sólo se percibe mediante metadatos externos generando un problema en el momento de transferir o transmitir los ficheros, dado que para su reconocimiento hay que controlar dos ficheros, el de los metadatos y el del objeto de datos que etiqueta.

Una investigación rápida de los formatos es posible acometerla mediante la Wikipedia (siempre con reservas). Si bien, sobre este asunto, se muestra como una eficaz fuente de información para formatos contemporáneos, dado que

las casas comerciales la emplean para promocionarlos. Obsérvese el ejemplo sobre el formato TIFF (Wikipedia, 2018) para comprender el grado de información que aportan.

No obstante, cuando trabajamos con formatos del “pasado”, es posible que dispongamos de formatos deficientemente identificados. De tal modo que ni siquiera podamos tener una idea de qué buscar en la Wikipedia o en otra herramienta de búsqueda. Para auxiliar en estos casos, han existido y coexisten iniciativas de gran valor. Destacamos por el alcance de las mismas, dos iniciativas de Registros. Estas son:

- 1) El Registro global de formatos digitales, cuyo primer proyecto lo lleva a cabo la Biblioteca de la Universidad de Harvard, en calidad de coordinadora, y que junto con la OCLC y la Fundación A. WE. Mellon, esta última como patrocinadora, desarrollan un *Registro global de formatos digitales*. En este gran proyecto de recogida de formatos participan también, la Biblioteca Nacional de Francia, la de Gran Bretaña, la *Library of Congress* y los Archivos Nacionales de Reino Unido y de Estados Unidos (NARA). Se trata de una de las primeras aplicaciones surgida para el control de los formatos, conocida como *Global Digital Formats Registry-GDFR*- (2018). Este proyecto ya ha concluido, derivando hacia un segundo proyecto de unificación de los formatos.
- 2) El Registro de formatos digitales unificados (UDFR), surgido en el 2012, ha sido el segundo gran proyecto acometido en este campo de los formatos, el cual aportaba las funcionalidades de identificar y caracterizar. En él confluyen las funcionalidades, de la base de datos –relacional- denominada PRONOM y la base de datos del GDFR. Se conformaba como una base de datos semántica mediante el modelo ontowiki. En la página del proyecto, es factible la consulta de información sobre su historia (Unified, 2016). Fue desarrollado por la *University of California Curation Center (UC3)* y la *California Digital Library (CDL)*, fundada por la [Library of Congress](#) como parte del *National Digital Information Infrastructure Preservation Program (NDIIPP)*.

A partir de dicho registro se genera la base de datos DROID que permite el reconocimiento de formatos desconocidos y trabaja en colaboración con la

herramienta PRONOM. Ambas iniciativas pasaremos a comentarlas en el siguiente enunciado.

Con independencia de investigar el formato mediante su extensión y número mágico o mediante la consulta de la Wikipedia, destacamos con un alcance más universal, dos iniciativas de calado: el GDFR y el UDFR. Por ende, las funcionalidades a destacar, en la primera mencionada (GDFR) son la posibilidad del reconocimiento e identificación del formato, frente a esta segunda iniciativa (UDFR), en la que se reconoce la identificación y caracterización, dado que estos fines los gestiona de modo semántico a partir de la herramienta ontowiki.

### **4.3 Aplicaciones específicas con prestaciones heterogéneas**

En este subepígrafe se enumeran y describen variadas herramientas con funcionalidades diversas para la investigación sobre formatos.

En primer lugar, destacamos aquel conjunto de aplicaciones disponibles de modo gratuito en la red, ampliamente utilizadas por informáticos y ciudadanía en general entre las que enumeramos las siguientes: a) *file extensions*; b) *files.com*; y c) *fileinfo.net*.

La herramienta, extensiones de ficheros (File-extensions, 2018) nos informa principalmente sobre las opciones de conversión de ficheros. Aporta alfabéticamente o por categorías de recursos digitales las extensiones. Contiene *miles de extensiones* de archivos y su base de datos continua creciendo. Aporta información de naturaleza normalizadora, pues informa sobre las herramientas de conversión de ficheros, permite identificar el formato siempre que en su base de datos se encuentre incorporada la extensión del fichero, dado que identifica precisamente a ésta. Por tanto, reconoce la extensión del fichero. Principalmente, identifica, normaliza y caracteriza algunos formatos. Incluso dispone de un listado de ficheros desconocidos. Facilita, además, información descriptiva sobre el software específico que es capaz de trabajar con el formato. Incluye características del fichero, descripciones detalladas, informando sobre el uso actual y la lista de programas que pueden abrir, ver, editar, convertir o reproducir archivo desconocido que busca.

La segunda herramienta mencionada, denominada “Fileinfo” (Fileinfo, 2018), facilita la identificación de formatos, mediante el mantenimiento de una base de datos de *extensiones* de ficheros. El enlace *fileinfo* aporta, como su nombre indica, información sobre el formato de fichero, el vendedor del software que lo interpreta, los programas o aplicaciones compatibles. Su principal función es la de identificar, aunque también describe sus propiedades (por lo que caracteriza al fichero). Es de muy sencilla consulta. Mediante ella, conocemos la extensión del formato, su descripción, apunta sobre su creador, señala el tipo de recurso digital y especifica las aplicaciones que lo abren. Esta herramienta dispone de vídeos ilustrativos sobre su potencial. Referenciamos un ejemplo<sup>3</sup>

El siguiente conjunto de herramientas responde a iniciativas efectuadas en unidades preservadoras, cuya característica principal es su relevante utilidad atendiendo a su finalidad creadora.

La primera de ellas, nos permite identificar la estructura del objeto digital. Por tanto, para aquel conjunto de recursos digitales en los que se desconoce el formato, esta herramienta presenta grandes oportunidades. Es conocida por sus siglas: DROID -*Digital Record Object IDentification* (National Archives of United Kingdom, 2018a). Se trata de una herramienta de código abierto y en software libre de identificación de formato de archivo desarrollada por el Archivo Nacional de Reino Unido. La función principal de DROID es la identificación exacta del formato de archivo, incluso si la extensión es incorrecta o se encuentra ausente. Aporta información amplia, hasta llegar incluso a la versión. Dispone de más de 1.400 formatos reconocidos\_y se encuentra en incremento constante. Además de la identificación del formato de fichero, desarrolla tres funciones, tal y como se expresa en su guía (National Archives of United Kingdom; 2017, p. 4) y que pasamos a enumerar:

- 1ª función: facilitar la comprensión del funcionamiento de la base de datos y su manejo.

---

<sup>3</sup> Disponible en: <https://www.youtube.com/watch?v=4PtGro50pfg&fe>.

- 2ª función: extraer otra información sobre los archivos que explora, tales como el tamaño del archivo, la fecha de la última modificación y la ruta del archivo. Esta información se presenta en un perfil que puede ser analizado en la interfaz gráfica de usuario de DROID (GUI) mediante filtrado o exportado a un archivo *CSV*.
- 3ª función: examina los ficheros dentro de archivos contenedores (tales como ficheros 'zip'), así como el propio archivo contenedor. Incluye hasta 18 metadatos específicos al formato (pp 8-9).

Esta herramienta DROID colabora con la iniciativa PRONOM estableciéndose un identificador único conocido como PUID, el cual facilitará su enlace a la iniciativa PRONOM, de utilidad para su caracterización, completando así la identificación. En definitiva, DROID es una herramienta de software que trabaja de modo colaborativo con la base de datos PRONOM. Identifica por lotes los formatos, escanea el objeto de datos de archivo y aporta un identificador único al formato para trabajarlo conjuntamente con *Pronom*. Como cualquier repositorio digital se ha diseñado para reconocer el formato y vincularlo, mediante el identificador.

Un segundo recurso a destacar por su importancia es el acrónimo conocido como XENA (*XML Electronic Normalising for Archives*) (National Archives of Australia, 2013). Al igual que DROID, es una aplicación de software libre y código abierto con licencia GNU pública general (GPL3), elaborado para la preservación digital y desarrollado por los Archivos Nacionales de Australia. Su finalidad es la de contribuir a la conservación a largo plazo de los recursos digitales. XENA es uno de los componentes que integran la Plataforma de Software de Preservación Digital (DPSP), en la que se incluye la validación de ficheros. Un paquete completo para la captura y conversión de ficheros en un repositorio. La aplicación australiana, realiza dos tareas básicas. Identifica y convierte a formatos abiertos para la conservación.

Los dos siguientes recursos que comentaremos proceden de la Biblioteca de la Universidad de Harvard. Nos referimos a las aplicaciones JHOVE2 (2013) y FITS.

La primera de ellas, incorpora múltiples funcionalidades; Es una herramienta que identifica (considerando su extensión y su número mágico); valida



(determina el nivel de conformidad del objeto digital mediante reglas normativas y sintácticas definidas por la especificación); caracteriza (informa sobre las propiedades intrínsecas de un objeto digital significativas para la preservación) y evalúa determinando su nivel de aceptabilidad para el propósito específico atendiendo a la política definida. Jhove2 facilita la captura –ingreso- su gestión, almacenamiento como PIA (paquete de información de archivo) y conservación en el repositorio digital. El PIT (paquete de información transferido) es identificado y chequeado para asegurar todas sus características. Una vez identificado el objeto y su información de representación es validado mediante la evaluación de la consistencia del formato de acuerdo al pliego de condiciones o convenio acordado y se mantienen los metadatos. Desempeña una función básica en un repositorio entre las entidades de ingreso y de gestión de datos. Cuando ya el fichero es reconocido como PIA. El problema mayor que presenta es que trabaja con una serie de familias de formatos y módulos estándar. No incluye gran cantidad aunque sí formatos muy vinculados a la conservación, tales como aquellas familias para ACII; GIF, JPEG, PDF, SGML, UTF, WAVE, XML, ZIP.

En su página web se dispone de numerosa información incluyendo: casos, módulos estándar, tutoriales, licencia Pública General GNU. Su versión actualizada es Jhove2, habiendo existido un proyecto anterior conocido como Jhove. Dispone de una guía para su aplicación

En definitiva, esta iniciativa de la Universidad de Harvard (Jhove2) incluye buena parte de las funcionalidades, por lo que ha de considerarse como una herramienta destacable. El único inconveniente que presenta es su limitación en cuanto al reducido número de formatos con los que trabaja.

La segunda iniciativa promovida por la Biblioteca de dicha Universidad es conocida bajo las siglas FITS (File Information, 2018). Se trata de una herramienta muy potente que permite identificar, validar y extraer metadatos.

Habiendo surgido como proyecto a ser aplicado en el repositorio digital de dicha Biblioteca, se encuentra actualmente en uso y dispone de manual y guía de usuario. Debe ser considerada como una plataforma que incorpora varias herramientas de acceso abierto y permite cierta interoperabilidad, actuando

como una envoltura o empaquetadora en un fichero XML. FITS está escrita en Java y es compatible con Java 1.7 o superior. Se utilizan herramientas externas preservadoras, las cuales son muy numerosas, tal y como se especifica seguidamente: [Herramienta ADL](#), [Apache Tika](#), [DROID](#), [Exiftool](#), FFIdent, [File Utility](#) (puerto de Windows), [Jhove](#), [Información de los medios](#), [Extractor de metadatos de la Biblioteca Nacional de Nueva Zelanda](#); [Información de audio OIS](#), [Información de archivo OIS](#), [Información XML de OIS](#).

En suma, sus funcionalidades más significativas son: identificar formato, validar y extraer metadatos. Esta última versión, como vemos resulta interesante porque facilita el control de los mismos. En esta línea, señalamos la siguiente iniciativa, igualmente, surgida bajo la mano de una biblioteca universitaria.

La quinta herramienta referenciada procede de la Biblioteca de la Universidad de Nueva Zelanda. Respondiendo a la función principal por la que destaca – extracción de metadatos- es conocida como *Metadata Extractor Tool* (National Library of New Zealand, 2007). Dicha aplicación se encuentra operativa desde el 2003, aunque en abierto, desde el 2007. Extracta los metadatos del fichero y permite su explotación en formato XML, de una gama de formatos de archivos para diferentes categorías de recursos: imágenes, documentos de office, audio, video, lenguaje de marcado y ficheros Web. Se orienta, evidentemente, a extraer aquellos metadatos útiles para la conservación. La aplicación ha sido realizada en JAVA y XML y se distribuye bajo la licencia Apache Pública. Dispone de una interfaz para Microsoft Windows y de una interfaz de comandos para UNIX. Dispone de una guía de usuario e información para instalación. Esta aplicación cumple la doble funcionalidad de identificar y extraer los metadatos de formato.

Finalmente, enumeramos una nueva aplicación, a la que ya hemos aludido en la presentación de DROID. Se trata de la herramienta PRONOM (National Archives of United Kingdom, 2018b), la cual ha de entenderse como un sistema de información en línea sobre los formatos de ficheros que en colaboración con DROID, facilita, una vez identificado el formato por esta última, sus propiedades, la viabilidad de su normalización hacia otra estructura de fichero o formato, los riesgos, información sobre creadores y mantenedores, información sobre software, sobre la documentación, etc.

Permite, además, su alimentación para añadir nuevos formatos e información al respecto: compresión, compatibilidad, derechos, riesgos, migración, metadatos, etc.

En la aplicación se define como un “*servicio de registro técnico que describe las dependencias técnicas de objetos digitales con el fin de apoyar la preservación a largo plazo*” Estas dependencias técnicas, que corresponden al concepto OAIS de información de representación, incluyen los formatos en los que se codifican los objetos, las herramientas de software que pueden ser necesarias para realizar acciones sobre dichos objetos (tales como creación, representación y migración) y el funcionamiento del sistema y hardware de esas herramientas.

En definitiva, dicha aplicación Identifica, caracteriza y normaliza. Es la herramienta por excelencia para la búsqueda de información sobre formatos en colaboración con la función identificadora de DROID. Ambas han sido elaboradas bajo la responsabilidad de la unidad preservadora de los Archivos Nacionales del Reino Unido. Al igual que DROID dispone de ayuda y una guía de usuario, la cual facilita su utilización.

Para finalizar, se ha elaborado una tabla en la que se reflejan las funcionalidades de las herramientas analizadas con el objetivo de mostrar una comparativa de las mismas sobre el conjunto de iniciativas analizadas.

Herramientas	Identifica	Valida	Caracteriza	Normaliza
<b>File-extensión</b>				
<b>Fileinfo.net</b>				
<b>DROID</b>				
<b>XENA</b>		metadatos		
<b>FITS</b>		metadatos		
<b>Extractor</b>		metadatos		
<b>Jhove2</b>		metadatos		
<b>PRONOM</b>		<i>Informa únicamente sobre metadatos</i>		

Tabla 5- comparativa de funcionalidades para las aplicaciones auxiliares sobre formatos. Fuente: elaboración propia

De este modo, se ofrece una visión general sobre las utilidades de los instrumentos presentados mediante su comparativa. Destacan las aplicaciones XENA y Jhove2 en cuanto que responden a todas las funcionalidades observadas. El tándem DROID y PRONOM, igualmente son aplicaciones muy recomendadas por su grado de vinculación y porque incluyen formatos antiguos. A su vez, un buen número de ellas validan metadatos del formato. Para búsquedas menos complejas, la herramienta *File-extensions*, *fileinfo.net* y la Wikipedia aportan valiosa información de utilidad para incrementar conocimiento de formatos. En el subapartado siguiente se amplía el número de iniciativas.

#### **4.4 Otras iniciativas de diferente naturaleza**

Dos son las iniciativas destacables con finalidades y procedencias diferentes. La primera, se corresponde con las acciones del Programa Nacional de Infraestructura de Información y Preservación (NDIIPP) (Library of Congress, 2017), del cual ya se ha aportado información. Resulta interesante retomar esa iniciativa para detallar que en el sitio, además de la información ya comentada, incorpora la percepción del preservador para cada categoría de recursos documentales, lo cual si cabe, convierte a este sitio en una ventana abierta a la percepción preservadora<sup>4</sup>, informado sobre características significativas de los formatos, subcategorías de características y preferencias de formatos.

La segunda iniciativa que se incluye es la referencia al sitio del preservador por excelencia, conocido bajo las siglas CORP -Community Owned Digital Preservation Tool Registry (CORP,2014). Se trata de un registro en el que colaboran múltiples entidades (DDC, NDSA entre otras) que incorpora el conjunto de herramientas útiles para la preservación, aportando información

---

<sup>4</sup> Se recomienda consultar el ejemplo. (LIBRARY OF CONGRESS, 2017) [en línea]. [Consulta:8 mayo 2018]. Disponible dicho ejemplo en: <http://www.loc.gov/preservation/digital/formats/content/text.shtml> .

sobre funciones, incluyendo la identificación, validación y migración de formatos. Una consulta al mismo, nos orienta sobre la cuantía de iniciativas existentes y consignadas en él, tales como: *NANITA*, una versión de *DROID6*; *FIDO*, herramienta de identificación basada en Pronom y similar a *DROID*; *Siegfried*, basada en Pronom, con prestaciones similares a *FIDO*; *Gvs-info*, sistema de archivo virtual con comando para imprimir información sobre ficheros promovido por Gnome; *FIDOO*, basada en Web para la identificación de formato; *Libmagic-dev*, identifica el número mágico del fichero; *LibsharedMime*, aporta el tipo MIME del fichero; aplicación *Apache TIKA*, referenciada al presentar FITS, identifica ficheros y extrae metadatos; *Obcount*, analiza ficheros de texto plano.

Incluye, además otras específicas en contextos determinados, por ejemplo, ficheros de Microsoft, como la herramienta *Officceparser.py* o la de *Web archives recovery*, específica para el indexado de contenido web, identificación de formatos web (ARC o WARC) e incluso especificación sobre algunos riesgos. Se menciona, igualmente, el analizador de ficheros *NARA-File and Metadata Harvester2* (NARA, 2015), el cual se encuentra alimentado por la Universidad Georgetown y el *Metadata Harvester*. Esta iniciativa facilita el análisis de los contenidos de un sistema de archivo o de cualquier unidad externa y genera estadísticas sobre los mismos considerando los directorios. Su utilidad facilita el *checksum* de los ficheros, aportando valores de suma de comprobación para garantizar la integridad de los archivos a nivel de bit, en los procesos de copiado o conversión. En esta línea, la COPTR aporta una relación de herramientas para la validación de archivos (CORP, 2014)<sup>5</sup> siendo computadas hasta 30. Igualmente, facilita el acceso a un listado de herramientas (CORP, 2014)<sup>6</sup> para la migración de formatos en las que registra 46 enlaces. Tal y como se deduce, el número de herramientas es relevante, aproximándose a 100.

## 5. Conclusiones

---

<sup>5</sup> El enlace directo de validación se encuentra disponible en:  
<http://coptr.digipres.org/Category:Validation>.

<sup>6</sup> El enlace directo a dichas herramientas se encuentra disponible en:  
[http://coptr.digipres.lrg/category:File\\_Format\\_Migration](http://coptr.digipres.lrg/category:File_Format_Migration).

El conocimiento del formato es vital para poder interpretar en el tiempo la información que contiene un objeto y sus propiedades. El proceso de decodificación de la información puede convertirse en un obstáculo a largo plazo ante el desconocimiento del formato. Por ende, su tipificación parece relevante en la preservación del contenido codificado digital. Máxime si acometemos un plan de preservación, aunque también resulta necesario identificar los formatos y determinar su viabilidad desde la creación y captura del recurso digital.

Se precisa disponer de la información sobre el formato mismo: sus reglas semánticas, sintácticas, extraer sus metadatos, su propietario, sus derechos, etc. La amplia diversidad de tipologías de formatos y su evolución derivan en un problema para la preservación, pues continuamente mudan las reglas. Por ende, tanto el incremento como la obsolescencia de formatos deben de ser resueltos desde el enfoque preservador con ciertas garantías de éxito.

Para ello, el conocimiento y uso de las iniciativas existentes auxilia en la identificación y control de los formatos. Estas aplicaciones disponen de funcionalidades diferenciadas para identificar, caracterizar, validar y normalizar.

El inventario de formato aporta garantías para identificar los formatos existentes de una organización. En coordinación con el catálogo de recursos y de dispositivos del entorno tecnológico ha de observarse como un arma eficaz para trabajar las obsolescencias y normalizar los formatos. La viabilidad de los mismos en el tiempo, sólo es posible mediante la selección de formatos y su reflejo mediante matriz. Principios y factores sobre los formatos contribuyen a reconocer criterios y formular indicadores para la selección. Algunos ejemplos han sido referenciados, no obstante, deberían considerarse las especificaciones y particularidades de cada centro. El resultado final, proveerá a la organización de un perfil de los formatos preferentes y aceptables para cada una de las categorías de recursos digitales existentes en ella. En este sentido, han sido detalladas las principales iniciativas, algunas de las cuales cumplen como se ha observado multiplicidad de funciones.

## **6. Recursos digitales y bibliográficos empleados**

ASOCIACIÓN ESPAÑOLA DE NORMALIZACIÓN (AENOR), 2011. *UNE-ISO 30301:2011 Información y Documentación. Sistemas de gestión para los documentos. Requisitos*. Madrid. Aenor.

AENOR, 2013. *UNE-ISO 13008:2013. Información y Documentación. Proceso de conversión y migración de documentos*. Madrid: Aenor.

BRAWN, A., 2008. *Digital Preservation Guidance Note 1: Selecting file formats for long-term preservation*, Londres: National Archives of United Kingdom. 10 p. [en línea]. [Consulta: 10 mayo 2018]. Disponible en: <http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>.

CORP –Community Owned Digital Preservation Tool Registry, 2014. *File Format Identification* [en línea] [Consulta: 2 mayo 2018] Disponible en: <http://coptr.digipres.org/Category:File Format Identification>.

CRUZ MUNDEZ, J.R. Y DÍEZ CARRERA, C., 2015. *Los costes de la preservación digital permanente*. Gijón: Trea. ISBN 978-84-9704-905-4.

DIGITAL PRESERATION COALITION, 2018. *Digital preservation Handbook. Files formats and standards*. [en línea] [Consulta: 5 mayo 2018] Disponible en: <https://www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards>.

ESPAÑA, 2017. Ministerio de Hacienda y de Administración Pública. *Código de la Administración Electrónica*. Madrid: Agencia Estatal del Boletín Oficial del Estado, (ed. Actualizada a 6/julio/2017). [en línea] [Consulta: 25 abril 2018]. Disponible en: URL: <http://www.boe.es/leg/codigos>.

FILE-extensions.org., 2018. *File-extensions.org* [en línea]. [Consulta: 2 abril 2018] Disponible en: <http://wwwfile-extensions.org>.

FILEINFO, 2018 *The file extensión data base*. [en línea]. [Consulta: 4 abril 2018]Disponible en: [www.fileinfo.net/](http://www.fileinfo.net/).

FILE Information Tool Set, 2018. [en línea]. [Consulta: 25 abril 2018] Disponible en: <http://projects.iq.harvard.edu/fits>.

GLOBAL Digital Format Registry (GDFR), 2016. [en línea]. [Consulta: 7 mayo 2018]. Disponible en: [https://library.harvard.edu/preservation/digital-preservation\\_gdfr.html](https://library.harvard.edu/preservation/digital-preservation_gdfr.html).

HARVARD LIBRARY DIGITAL PRESERVATION, 2017. *Format Matrix Template*. [en línea]. [Consulta: 10 mayo 2018]. Disponible en: [https://docs.google.com/spreadsheets/d/1buM2XZtkc09kUtUo0W5s0lt4IK\\_6LALF6VooCJDdQZ0/edit#gid=1603236](https://docs.google.com/spreadsheets/d/1buM2XZtkc09kUtUo0W5s0lt4IK_6LALF6VooCJDdQZ0/edit#gid=1603236).

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION- ISO, 2017. *ISO 28500:2017. Information and documentation. WARC file format*. Génève. ISO.

ISO TC46/SC11WG7 Digital Records Preservation, 2010. Preservación de documentos digitales. Guía “Cómo empezar. Aenor (trad.) [en línea]. [Consulta: 10 mayo 2018]. Disponible en: <https://www.arxiv.org/index.php/documents/documentacion/normativa-tecnica-1/514-tec-isotc46sc11-wg7-preserv-docs-digit-es-1/file>.

JHOVE2, 2013. [en línea]. [Consulta: 25 abril 2018]. Disponible en: <https://bitbucket.org/jhove2/main/wiki/Home>.

LIBRARY OF CONGRESS, 2017. *Sustainability of Digital Formats: Planning for Library of Congress Collections*. [en línea]. [Consulta: 8 mayo 2018] Disponible en: [https://www.loc.gov/preservation/digital/formats/intro/format\\_eval\\_rel.shtml](https://www.loc.gov/preservation/digital/formats/intro/format_eval_rel.shtml).

LIBRARY OF CONGRESS, 2018a. *Sustainability of digital formats Planning for Library of Congress collections*. [en línea]. [Consulta: 7 mayo, 2018] Disponible en: <http://www.loc.gov/preservation/digital/formats/sustain/sustain.shtml>.



LIBRARY OF CONGRESS, 2018b *Geospatial content. Quality and functionality factors*. [Consulta: 7 mayo, 2018]. Disponible en: [https://www.loc.gov/preservation/digital/formats/content/gis\\_quality.shtml](https://www.loc.gov/preservation/digital/formats/content/gis_quality.shtml)

LIBRARY OF CONGRESS, 2018c. *Recommended Formats Statement*. [Consulta: 9 mayo 2018] Disponible en: <https://www.loc.gov/preservation/resources/rfs/TOC.html>.

NATIONAL ARCHIVES AND RECORDS ADMINISTRATION (NARA), 2015.  
*NARA. File Analyzer and Metadata Harvester*. [en línea] [Consulta: 2 mayo 2018].  
Disponible en: [http://coptr.digipres.org/NARA File Analyzer and Metadata Harvester](http://coptr.digipres.org/NARA_File_Analyzer_and_Metadata_Harvester).

NATIONAL ARCHIVES AND RECORDS ADMINISTRATION (NARA), 2018.  
*Appendix A: Tables of file formats*. [en línea]. [Consulta: 9 mayo 2018] Disponible en: <http://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html#scannedtext>.

NATIONAL ARCHIVES OF AUSTRALIA, 2013. *XENA. Software for Digital Preservation*. [en línea]. [Consulta: 2 abril 2018]. Disponible en: <http://xena.sourceforge.net/>.

NATIONAL ARCHIVES OF AUSTRALIA, 2018. *Digital Continuity 2020*. [en línea] [Consulta: 9 mayo 2018] Disponible en: [http://www.naa.gov.au/Images/Preservation\\_File\\_0Formats\\_tcm16-99797.pdf](http://www.naa.gov.au/Images/Preservation_File_0Formats_tcm16-99797.pdf).

NATIONAL ARCHIVES OF UNITED KINGDOM, 2011 *Evaluating your file format*. Versión 1.2. Disponible en: [www.nationalarchives.gov.uk/digitalcontinuity](http://www.nationalarchives.gov.uk/digitalcontinuity) [data de consulta: 10 mayo 2018]

NATIONAL ARCHIVES OF UNITED KINGDOM, 2017. *DROID. User guide*. [en línea]. [Consulta: 10 mayo 2018]. Disponible en: <http://www.nationalarchives.gov.uk/documents/information-management/droid-user-guide.pdf>.

NATIONAL ARCHIVES OF UNITED KINGDOM, 2018a *DROID file-format identification tool*. [en línea]. [Consulta: 2 abril 2018] Disponible en: <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>.

NATIONAL ARCHIVES OF UNITED KINGDOM, 2018b. *The technical Registry PRONOM*. [en línea] [Consulta: 3 mayo 2018] Disponible en: : <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

NATIONAL LIBRARY OF NEW ZEALAND, 2007 *Metadata Extraction Tool*. [en línea] [Consulta 2 mayo 2018]. Disponible en: <http://meta-extractor.sourceforge.net/>.

QUEENSLAND GOVERNEMENT, 2018. *Digital Records Formats*. [en línea]. [Consulta 10 mayo 2018]. Disponible en: <https://www.forgov.qld.gov.au/digital-record-formats> .

ROSENTHAL, D, 2015. *Format Obsolescence: the Prostate Cancer of Preservation*. [en línea] [Consulta: 11 mayo 2018]. Disponible en: <https://blog.dshr.org/2007/05/format-obsolescence-prostate-cancer-of.html>.

UNIFIED Digital Format Registry (UDFR), 2016. [en línea]. [Consulta: 7 mayo 2018]. Disponible en: <http://udfr.org/project/>.

VOUTSSÁS, J., 2013 *Cómo preservar mi patrimonio digital personal México: Instituto de Investigaciones bibliotecológicas y de información*. ISBN: 978-607-02-4122-2 [en línea] [Consulta: 10 mayo 2018] Disponible en: [http://132.248.242.6/~publica/archivos/libros/como\\_preservar\\_patrimonio\\_dig\\_pers.pdf](http://132.248.242.6/~publica/archivos/libros/como_preservar_patrimonio_dig_pers.pdf).

WIKIPEDIA. La enciclopedia libre, 2018. *TIFF*, [en línea]. [Consulta: 7 mayo 2018]. Disponible en: <https://es.wikipedia.org/wiki/TIFF>