

KULTURA 2.0

CUADERNOS DE FORMACIÓN

2

FORMATOS DE DIFUSIÓN y FORMATOS
DE PRESERVACIÓN DE CONTENIDOS DIGITALES

JULIO 2011

FORMATOS DE DIFUSIÓN Y FORMATOS DE PRESERVACIÓN DE CONTENIDOS DIGITALES JULIO 2011

En el marco de la iniciativa kultura 2.0, el Observatorio Vasco de la Cultura inicia la elaboración y difusión de la colección Kultura 2.0-Cuadernos de Formación, destinada a difundir el conocimiento, la innovación y las buenas prácticas entre las empresas, principalmente pymes, y los profesionales del sector de la cultura de la CAE.

La finalidad es responder a las necesidades de información y formación que plantean estas empresas y los profesionales para adaptarse a la evolución, las nuevas herramientas, prestaciones y servicios en el ámbito de la denominada Kultura 2.0.

Estas necesidades se han identificado a través de grupos de trabajo en los que han participado activamente agentes del sector en la CAE. Para complementar y dar una mayor dinamicidad a este diagnóstico inicial de necesidades, se contempla recopilar sugerencias y propuestas de temáticas que puedan abordarse en próximos números.

Esperamos con esta iniciativa facilitar a las empresas y profesionales que trabajan en las actividades culturales metodologías y herramientas que les permitan ir adaptándose a los avances y oportunidades de la web 2.0.

Propuesta de temas para nuevos cuadernos

¿Qué otros temas sobre kultura 2.0 consideras podríamos tratar en próximos cuadernos de trabajo? (concretar lo mejor posible)

Envíanos tus sugerencias al mail: kulturabehatokia@ej-gv.es

colección

1. Visibilidad en la Red 2.0: Pautas para Mejorar y Monitorizar Nuestra Presencia en la Red - Marzo 2011
2. Formatos de Difusión y Formatos de Preservación de Contenidos Digitales - Julio 2011

Índice

1.	PRESENTACIÓN	4
2.	CIFRAS PARA SITUAR LA DIGITALIZACIÓN	5
3.	DIGITALIZACIÓN: CONCEPTO Y FINALIDAD	8
4.	CRITERIOS PARA LA SELECCIÓN DE LOS FORMATOS DIGITALES	11
5.	CARACTERÍSTICAS PARA DIGITALIZAR SEGÚN EL TIPO DE DOCUMENTO	13
6.	ALGUNAS RECOMENDACIONES FINALES	19
7.	RECURSOS COMPLEMENTARIOS	23
	Conceptos básicos	24
	Instituciones internacionales de referencia en digitalización	26
	Selección de proyectos de interés y buenas prácticas	26

1. PRESENTACIÓN

En los últimos años, la producción de contenidos digitales ha experimentado un crecimiento progresivo vinculado principalmente a los avances tecnológicos relacionados, en gran medida, con el crecimiento de la capacidad de almacenamiento, el desarrollo de la banda ancha y la velocidad de acceso a Internet, o la aparición de nuevos dispositivos electrónicos; factores que han incrementado las posibilidades de creación, distribución y consumo de dichos contenidos.

Este escenario ha generado un punto de encuentro interesante entre las industrias creativas y los sectores más relacionados con el ámbito tecnológico, y supone una oportunidad estratégica para el sector de la cultura, que se ha convertido en uno de los principales proveedores de contenidos del mercado digital. Esta confluencia no queda exenta, en todo caso, de algunas consecuencias nocivas para los modelos de negocio más tradicionales, así como ciertas desavenencias en la definición y explotación de los nuevos modelos de negocio.

Son amplias las expectativas que, desde un punto de vista cuantitativo, puede aportar la vertiente digital de la cultura como sector emergente generador de empleo y crecimiento económico. Pero, además de la potencial trascendencia en el plano socioeconómico, no debemos olvidar la relevancia de la digitalización desde una perspectiva antropológica, y sobre todo en el ambicioso objetivo de preservar y divulgar el patrimonio cultural y la memoria colectiva, motivos que han hecho de la digitalización uno de los ejes de la [política cultural europea](#)¹.

Este Cuaderno de Formación tiene por finalidad introducir al usuario en el complejo mundo de la digitalización. Trataremos de identificar y describir los formatos digitales de uso común y las tendencias y recomendaciones técnicas observadas respecto a su aplicación en actividades de preservación y difusión, centrando el foco de atención en los principales tipos de contenidos digitales: texto, imagen, audio y vídeo, así como en los formatos no ligados a soportes físicos o medios de almacenamiento determinados.

No es nuestro objetivo realizar un inventario de todos los formatos digitales —sobre todo teniendo en cuenta su rapidísima evolución y actualización—, ni profundizar sobre qué y cómo digitalizar. Nuestra intención es proporcionar al lector unos primeros flashes sobre los formatos digitales y ofrecerle referencias técnicas que le orienten a un primer nivel sobre aspectos que sean de su interés.

¹ Puede consultarse el documento «[Una Agenda Digital para Europa](#)» Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las regiones. Bruselas, 26.8.2010 COM (2010) 245 final/2. La finalidad genérica de la Agenda Digital es obtener los beneficios económicos y sociales sostenibles que pueden derivar de un mercado único digital basado en una Internet rápida y ultrarrápida y en unas aplicaciones interoperables.

2. CIFRAS PARA SITUAR LA DIGITALIZACIÓN

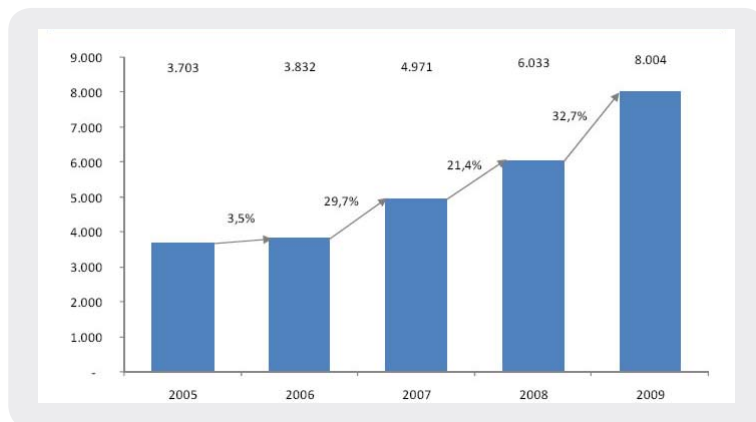
La evolución de los contenidos digitales puede analizarse desde puntos de vista diversos: económicos, sociales, culturales, tecnológicos, etc. Es difícil resumir en pocas páginas la extensa información cuantitativa y cualitativa que actualmente se difunde sobre los contenidos digitales y su desarrollo progresivo. En todo caso, para disponer de una panorámica sucinta sobre la situación actual y las perspectivas de esta actividad, se apuntan a continuación algunos datos de interés.

Evolución de la industria de contenidos digitales¹

- Analizando la evolución de la industria de contenidos digitales en los últimos años, observamos que el incremento ha sido espectacular. Así por ejemplo, **en 2009 superó en España una facturación de 8.004 millones de euros**; cifra que en relación al año anterior, supone un incremento del 32,7%. Respecto a la facturación total de la industria de los contenidos y servicios audiovisuales, los contenidos específicamente digitales constituyen el 45% de esta cifra.
- Las actividades relacionadas con el audiovisual y el cine (40,8%), así como el vídeo y los programas de televisión (29,2%), son las que contribuyen en gran medida a la facturación de los contenidos digitales.
- Se prevé que las Tecnologías de la Información y Comunicaciones (TIC) pasen a convertirse en los próximos quince años en el segundo mercado con mayor contribución al Producto Interior Bruto (PIB), teniendo a la industria de los contenidos digitales a la cabeza del crecimiento.

¹ Fuente: [ONTSI. Informe Anual de los Contenidos Digitales en España 2010](#)

Figura 1. Facturación de Contenidos Digitales: evolución 2005-2009. (millones de euros)



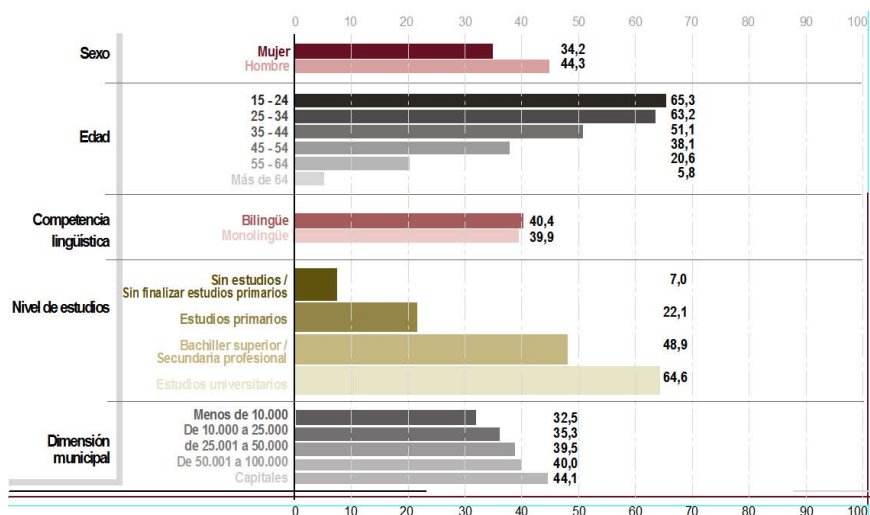
Fuente: [ONTSI, Informe Anual de los Contenidos Digitales en España 2010](#)

Hábitos culturales en la era digital

- Sumado a este espectacular despegue de la industria de los contenidos digitales, es necesario reflejar la **importancia que día a día va cobrando «lo digital» en los hábitos de consumo culturales¹** de la ciudadanía. Reseñar como ejemplo la fuerza que va adquiriendo poco a poco Internet como medio de consumo cultural en la Comunidad Autónoma de Euskadi, y sobre todo fijarnos en el perfil de los usuarios más cercano a esta vía de consumo.
- El uso de Internet para el consumo cultural se encuentra **más extendido entre los hombres, la población más joven** (sobre todo 15-24), y con un nivel de estudios **universitario**.

¹ Estadística de Hábitos, Prácticas y Consumo en Cultura CAE 2007-2008. Observatorio Vasco de la Cultura.

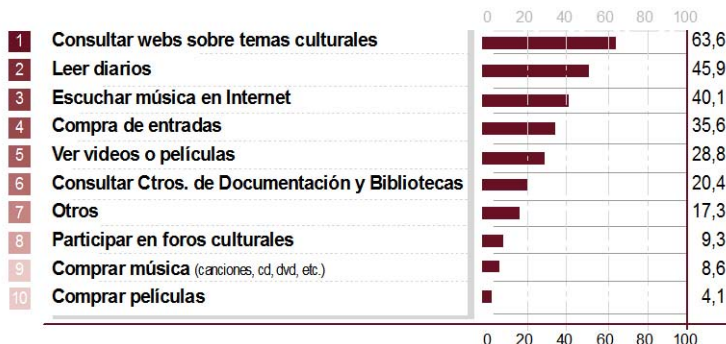
Figura 2. Personas que utilizan Internet para realizar algún tipo de consumo cultural por sexo, edad, competencia lingüística, nivel de estudios y dimensión municipal. (%)



Fuente. *Estadística de Hábitos, Prácticas y Consumo en Cultura CAE 2007-2008*. Observatorio Vasco de la Cultura.

- **Casi 4 de cada 10 personas utiliza la red para realizar algún tipo de consumo cultural.** El 46% de estas personas lee las ediciones digitales de los diarios y el 40% escucha música en Internet. Casi el 30% ve películas o vídeos a través de la red y más del 20% consulta Centros de Documentación y Bibliotecas.

Figura 3. Personas según su acceso a contenidos culturales en Internet (%)



Fuente. *Estadística de Hábitos, Prácticas y Consumo en Cultura CAE 2007-2008*. Observatorio Vasco de la Cultura.
 Base: Personas que utilizan Internet para realizar algún tipo de consumo cultural, 39,1% de la población

- Desde otra óptica es interesante el dato que hace mención a la **comprensión lectora en formato digital**. Según el último [Informe PISA 2009](#), que presenta resultados sobre la Evaluación de la Lectura de Textos Electrónicos (ERA, en sus siglas en inglés)¹ en jóvenes de 15 años, España se sitúa con 475 puntos en una baja posición, 24 puntos por debajo de la media de los países de la OCDE que han participado en el estudio. La prueba tiene como objetivo medir el rendimiento de los alumnos en el acceso, comprensión, valoración e integración de textos electrónicos en un espectro variado de contextos y actividades de lectura.

¹ La muestra sobre habilidades de lectura digital en España estaba compuesta únicamente por unos 2.300 alumnos de 170 centros, por ello, los datos a nivel autonómico no tienen valor estadístico.

- Finalmente, y respecto a los soportes, señalar que entre los lectores en soporte digital en España, en torno al 48% lo hace en el ordenador¹, un 7% utiliza el móvil o la agenda electrónica para leer contenidos en soporte digital, y un 1,3% lee en un E-reader.

kultura2.0

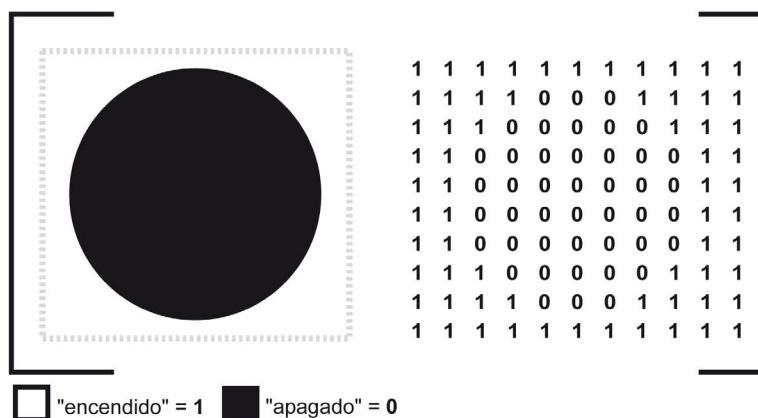
3. DIGITALIZACIÓN: CONCEPTO Y FINALIDAD

El concepto de digitalización

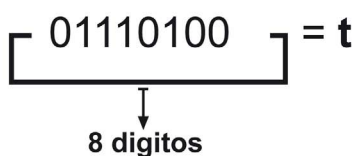
Digitalizar, definido de forma muy simple, es traducir información a un formato que puedan entender los ordenadores.

Se trata de una operación mediante la cual se convierte un texto, una imagen, un sonido, etc., en una serie de códigos binarios (un ordenador sólo está capacitado para manejar los valores 1 y 0) que representan cada uno de los puntos de su estructura, como puede verse en el ejemplo siguiente, permitiendo que esta información sea almacenada en el ordenador.

ejemplo imagen



ejemplo texto



¹ Estadística de hábitos de lectura y compra de libros en España 2010. Federación de gremios de Editores de España.

Al digitalizar convertimos contenidos en series de bits y los almacenamos en soportes electrónicos tangibles (CDR, DVD, USB, etc.) o intangibles (memoria del ordenador).

La digitalización permite socializar la información, y ha posibilitado como nunca antes la movilización de la información, ampliando su accesibilidad y la gestión del conocimiento.

A la hora de digitalizar es preciso tener en cuenta...

- Evidentemente, el **tipo de contenido** con el que contamos: si es un texto, una imagen, un sonido, un vídeo...
- Las **características específicas de cada documento**: así por ejemplo, no se siguen las mismas especificaciones para digitalizar una foto en blanco y negro o una foto en color,
- El **origen del formato**: es decir si es analógico o digital, y
- La **finalidad del archivo digital y su ubicación en la cadena de valor**: formato inicial de creación, formato intermedio para edición, formato final para preservación o difusión.

Objetivos y razones de la digitalización

Los objetivos de la digitalización pueden ser diversos pero simplificando el análisis podemos identificar dos tipos de finalidad: la **preservación y la difusión de los contenidos**.

En general son las instituciones culturales, como por ejemplo bibliotecas, archivos o museos, las que han asumido la responsabilidad de la preservación a largo plazo de los contenidos culturales en el marco de su función social, aunque esta labor también afecta a la industria.

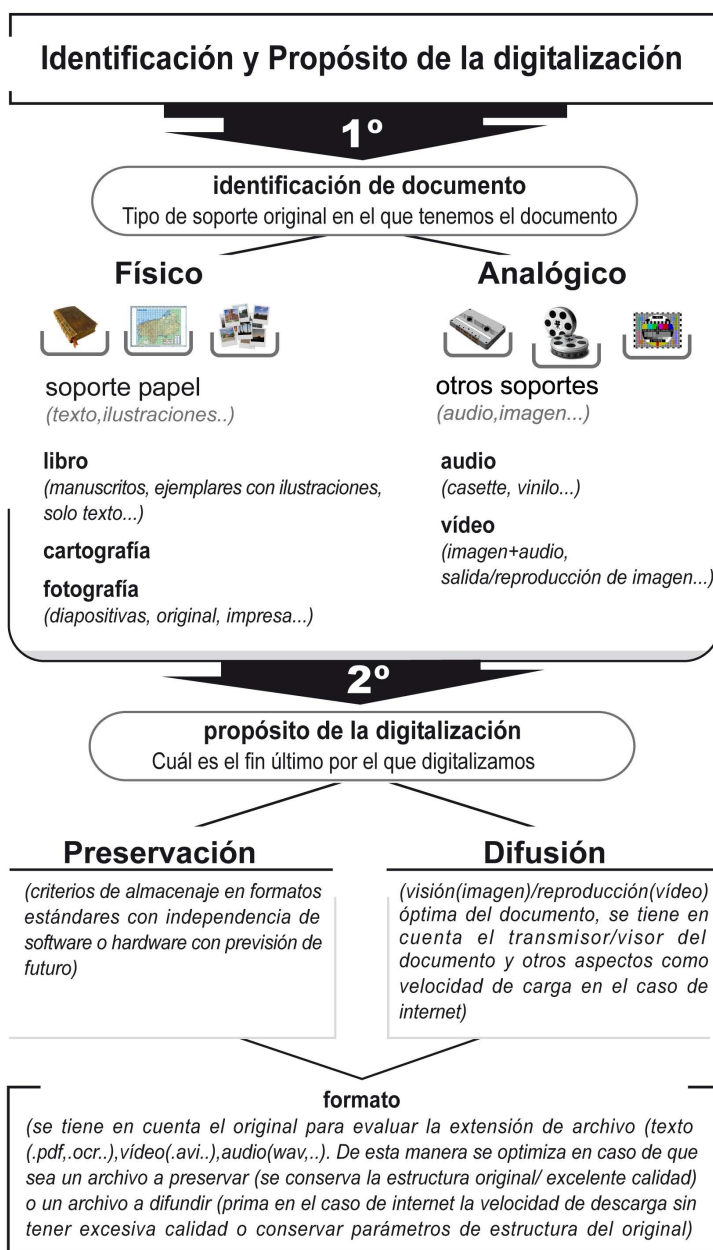
La difusión de los contenidos, ya sea con fines comerciales o para mejorar la accesibilidad de la ciudadanía a los mismos, es una cuestión que afecta de igual modo al ámbito público y a la industria.

Las razones por las que se digitalizan los contenidos pueden resumirse de la forma siguiente:

- Incrementar el acceso: permitir el acceso del mayor número de documentos al mayor número de posibles usuarios.

- Preservar: necesidad de anticipar, prevenir, detener, retardar un documento (texto, sonido, imagen, vídeo).
- Disminuir la manipulación y el uso de materiales originales que puedan ser frágiles, únicos o muy utilizados y crear una «copia de seguridad» para documentos en mal estado.
- Facilitar nuevas formas de acceso y uso.

El gráfico que se presenta a continuación sintetiza las dos premisas básicas del proceso a la hora de digitalizar: **qué tipo de documento tenemos y para qué deseamos digitalizarlo.**



4. CRITERIOS PARA LA SELECCIÓN DE LOS FORMATOS DIGITALES

Un **formato digital** podemos definirlo, de manera sencilla y alejándonos de tecnicismos, como el sistema de codificación de la información para su posterior almacenamiento o tratamiento en un soporte informático (documentos trabajados desde los programas del ordenador) o digital (captura de imágenes, por ejemplo, desde un escáner, una cámara de fotos digital).

En la práctica, el desarrollo de contenidos digitales ha derivado en multitud de formatos (se distinguen por el contenido del documento; así pues si un archivo es de texto el formato «extensión de archivo» es .doc; si es imagen es .tiff .jpeg; si es audio .mp3, etc.), que pueden estar diseñados para objetivos diversos:

- almacenar un solo tipo de contenido sin ninguna codificación adicional;
- incorporar especificaciones para codificar la información (principalmente para su compresión, transmisión o cifrado); o
- combinar y sincronizar varios tipos de contenido en un solo archivo. Por ejemplo en archivos multimedia como .mpeg o .AVI se almacenan pistas de audio, vídeo, subtítulos, metadatos, etc.

A la hora de seleccionar el formato más adecuado para un documento, se aconseja tener en cuenta dos categorías básicas de factores, recomendados por instituciones internacionales de referencia (ver apartado 7 de recursos complementarios), y que se relacionan con la finalidad de la propia digitalización (preservación/difusión) como queda reflejado en el gráfico siguiente.



- a. **Factores de sostenibilidad.** Cuando la finalidad de un documento (por ejemplo un manuscrito de Leonardo Da Vinci) es su preservación, existen diversos factores transversales a todos los tipos de contenido que deben considerarse para facilitar la permanencia de los documentos en el tiempo. Estos factores resultan determinantes para hacer más viable la [migración](#) de estos contenidos y minimizar los costes de preservación. Son los denominados [factores de sostenibilidad](#), entre los que destacan la interoperabilidad de los formatos o la independencia de los mismos de un software o hardware concreto (ver relación de factores establecidos por entidades de referencia internacionales —tabla A).

TABLA A-Relación de factores de sostenibilidad a tener en cuenta en la selección de los formatos establecidos por entidades internacionales de referencia

- **Divulgación/Transparencia:** hace mención al grado en el que las especificaciones técnicas de los formatos son accesibles (es decir son de código abierto).
- **Apertura:** se refiere a la dependencia de un formato determinado a una patente. En el caso de las entidades dedicadas a la preservación y/o difusión del patrimonio cultural, en general, se tiende a utilizar formatos abiertos, es decir formatos de archivo que no se hallan sujetos a patentes o derechos de autor. En el caso de las industrias culturales, existe un vínculo creciente entre las plataformas distribuidoras de los contenidos y los fabricantes de software y dispositivos (Apple, Amazon, etc.), que pueden imponer formatos propietarios (protegidos por patentes o derechos de autor) siguiendo criterios comerciales.
- **Dependencia/Interoperabilidad:** se trata del grado de dependencia de un formato determinado respecto a un hardware, a un software o a un sistema operativo específico.
- **Estabilidad/Compatibilidad:** grado en el que un formato mantiene su funcionalidad e integridad con versiones anteriores o posteriores.
- **Aceptación:** hace referencia al grado de utilización de los formatos por parte de los creadores, distribuidores y usuarios de los recursos.
- **Estandarización:** adecuación formal a los procesos o especificaciones establecidos por un organismo de normalización (ISO, NISO, W3C...) con el objetivo de garantizar la calidad de los archivos y su interoperabilidad.
- **Mecanismos de protección técnica:** los mecanismos de protección técnica como por ejemplo el cifrado, utilizados habitualmente para proteger la propiedad intelectual, no deben dificultar la recuperación de datos, migración de los contenidos o su adaptación a nuevas necesidades derivadas de la evolución tecnológica.

- b. **Factores de calidad y funcionalidad.** La elección del formato se fundamenta básicamente en su adecuación a las características del contenido y a las expectativas del usuario. En este sentido, las características intrínsecas de los formatos, generalmente específicas para un tipo de contenido —por ejemplo la resolución en el caso de las imágenes o la velocidad de transferencia en el vídeo—, constituyen los **factores vinculados a la calidad y funcionalidad** de los formatos.

5. CARACTERÍSTICAS PARA DIGITALIZAR SEGÚN EL TIPO DE DOCUMENTO

En este apartado presentamos las características a considerar cuando se digitaliza texto, imagen, audio, o vídeo, bien sea para la preservación y/o la difusión.

Digitalizar Texto

El primer indicador a estimar para la selección del formato es la tipología de documento con el que tenemos que trabajar dada la diversidad existente, por ejemplo:

- textos cortos que presentan una estructura simple;
- documentos en los que la estructura y sus características de navegación son claves para su comprensión;
- documentos en los que prima el diseño del contenido o contienen elementos visuales de mayor relevancia que el propio texto;
- documentos que combinan texto con otros elementos visuales como diagramas o fórmulas matemáticas (informes técnicos...);
- documentos accesibles para personas con discapacidades visuales; etc.

Cuando deseamos **digitalizar un texto con la intención de preservar**, tendremos en cuenta la estructura original (tipografía utilizada, color de texto, sangrado, tamaño de texto...): si trabajamos con un manuscrito primará la estructura y la calidad de detalle, y la imagen deberá digitalizarse con una alta resolución; si se transcribe el texto del manuscrito primará la información contenida en el mismo.

Cuando nuestro objetivo sea **difundir un texto**, si se desea mantener la estructura del mismo se digitalizará como imagen (el texto no es carácter). Puede digitalizarse un texto sin mantener la estructura a través de hardware para identificar automáticamente símbolos o caracteres del alfabeto ([reconocimiento óptico de caracteres -OCR](#)).



imagen extraida del archivo digital de la web oficial de Leonardo da Vinci. "Códice Atlántico" <http://www.leonardodigitale.com>

(1) imagen extraida del archivo digital de la web oficial de Leonardo da Vinci. "Traitté de la peinture, 1651" <http://www.leonardodigitale.com>

Digitalizar Imagen

La tipología de imágenes es también diversa y será el primer input a considerar al digitalizar éstas, así por ejemplo:

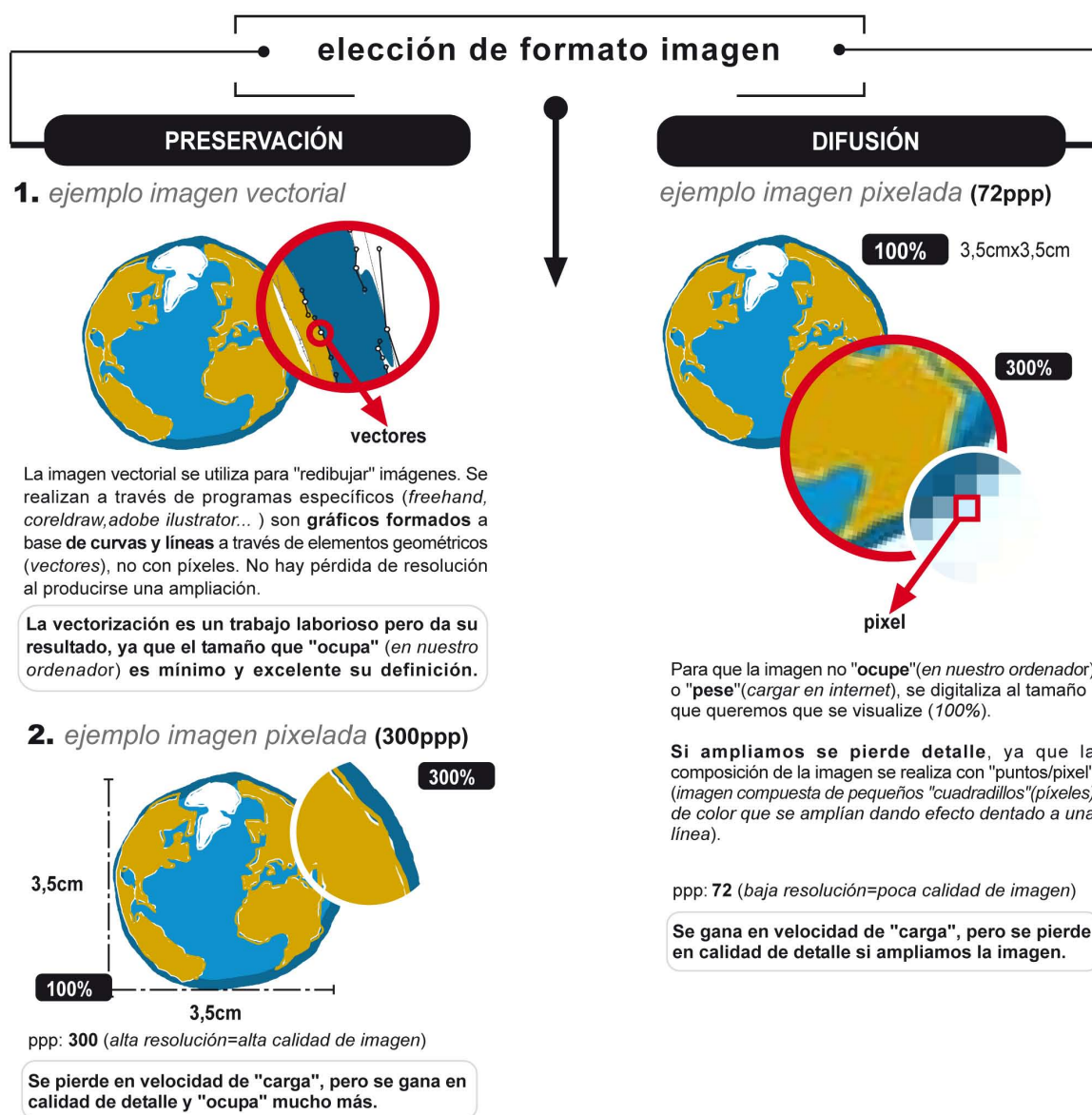
- imágenes en las que resulta esencial reflejar las características del documento original, por ejemplo, en expresiones artísticas como un cuadro o una fotografía;
- documentos donde la representación espacial es lo más importante y pierden valor otros aspectos como el color (mapas, planos...);
- imágenes con valor documental pero sin un gran valor artístico, etc.

No obstante, cuando vayamos a trabajar con imágenes es preciso distinguir **dos tipos básicos de formatos**:

- **gráficos o imágenes pixeladas o de mapa de bits (raster images):** las imágenes están representadas por píxeles. Es el tipo de formato que se utiliza comúnmente para captar imágenes con cámaras digitales y escáneres.

- imágenes vectoriales:** están definidas matemáticamente y utilizan formas geométricas para representar las imágenes. En este caso la información que almacenan no hace referencia a los píxeles. Se utilizan principalmente en actividades como el diseño gráfico, ingeniería, etc. El tamaño de los ficheros es muy pequeño.

En el gráfico siguiente se representan las diferencias más significativas de ambos formatos.



Las dos características básicas que determinan la calidad de la imagen, principalmente en las imágenes pixeladas, son la resolución y la profundidad de color o profundidad del bit.

La **resolución** es la cantidad de puntos que componen una imagen, es decir, la capacidad de distinguir los detalles espaciales más finos. Generalmente las recomendaciones encontradas respecto a la resolución de las imágenes se refieren a la resolución de impresión, es decir, a la cantidad de píxeles en relación con el tamaño de la imagen. Habitualmente se expresa en píxeles por pulgada (ppp o ppi en inglés). Lo importante es que **a mayor número de píxeles, mayor calidad de la imagen y mayor peso del archivo.**

La **profundidad de color o profundidad del bit**, se refiere al número de bits utilizados para describir el color de cada píxel. **Cuanto mayor sea la profundidad de bits mayor será la cantidad de tonos representados.**

Profundidad de Color	Nº de colores visibles
1 bit (monocromo)	2 tonos
4 bit	16 tonos
8 bit (indexed colour)	256 tonos
24 bit (true color)	16.777.216 tonos

Digitalizar Audio

En el caso de contenidos de audio, hay también distintas tipologías de contenidos digitales, por ejemplo, audio surround (multicanal), mono o estéreo, streaming, codificación digital de archivos no musicales (efectos de sonido, voz...), etc. Se distinguen **dos categorías elementales de formatos**:

- Por un lado, aquellos que recogen en forma de datos el sonido creado directamente a través de dispositivos electrónicos como ordenadores, sintetizadores, etc. para que después sean interpretados por el software utilizado para su reproducción.
- Por otro, los formatos más comunes y que interesan en este caso, los utilizados en las grabaciones musicales, de voz o de cualquier otro tipo, como por ejemplo el WAV, que almacena muestras de la onda de sonido para su posterior reproducción (formatos waveform).

En contenidos de audio son dos las características principales a considerar:

- **Frecuencia de muestreo:** número de muestras que se toman de la onda de sonido en un período de tiempo, se expresa en Khz.
- **Profundidad de bits:** número de bits utilizados para representar la onda sonora.

A mayor frecuencia de muestreo y profundidad de bits, mayor calidad del sonido. Por ejemplo, para obtener una calidad de CD de audio, la grabación se realiza a 44,1 Khz y 16 bits.

Como se ha representado en el gráfico siguiente con una pista ficticia, en la preservación de audio se recogen todas las frecuencias del archivo. En cambio, cuando sólo se va a difundir, se eliminan espacios vacíos y algunas frecuencias del archivo.



Digitalizar Vídeo

Cuando nos referimos a formatos de vídeo, generalmente estamos hablando de formatos contenedores o formatos de compresión de información audiovisual. Algunos documentos también se refieren a los formatos de vídeo como formatos multimedia, ya que pueden contener varios tipos de información: pistas de vídeo, pistas audio, en ocasiones texto o incluso imágenes incrustadas para los subtítulos y los metadatos necesarios para su sincronización.

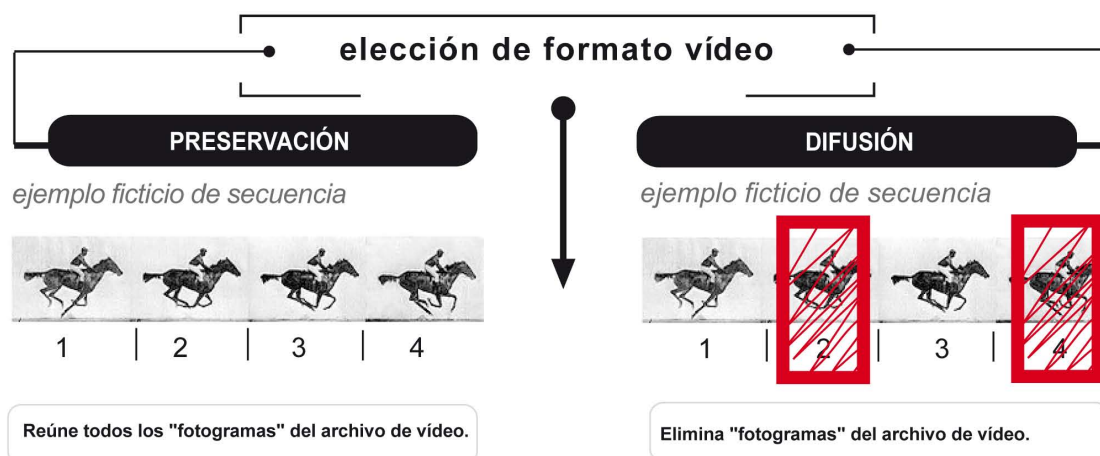
Aunque en el resto de contenidos (texto, imagen, audio) los formatos y estándares siguen evolucionando se observa un mayor consenso respecto a los formatos y especificaciones técnicas recomendadas. En el caso de los contenidos audiovisuales resulta más complejo establecer unas directrices de uso común, especialmente en el ámbito de la preservación.

La principal dificultad estriba en la amplia gama de formatos audiovisuales existentes y su mayor complejidad técnica, ya que contienen diversos tipos de información (audio, imagen...) y sus especificaciones suelen estar diseñadas para productos y canales de difusión específicos, como por ejemplo: formatos profesionales de grabación y edición, formatos profesionales para la distribución en salas de cine, formatos para la distribución en DVD, formatos para la difusión en Internet, etc.

Las recomendaciones técnicas elementales, se refieren a la imagen de vídeo digital aunque las especificaciones más complejas incluyen el audio, sistemas de compresión, etc. La imagen de vídeo, está compuesta por una sucesión de fotogramas, cada uno de los cuales contiene una imagen pixelada, y su calidad está determinada esencialmente por los métodos de captura y almacenamiento utilizados, que pueden fijar los aspectos clave que se describen a continuación:

- **Resolución de la imagen:** definida por el número de píxeles («puntos») que componen la imagen.
- **Velocidad de la imagen (*frame rate*):** número de imágenes estáticas o fotogramas por segundo, se expresa en fps (*frames per second*). Existen distintos estándares para cine (24 fps) y los distintos sistemas de televisión, etc.
- **Tasa de bits (*bit rate*):** velocidad de transferencia de los datos contenidos en el archivo audiovisual. Se expresa en bits por segundo (bps) y determina la nitidez de la imagen: a mayor velocidad más información puede contener la imagen.
- **Profundidad de color:** al igual que en las imágenes estáticas, se refiere al número de bits utilizados para describir el color de cada píxel.
- **Relación de aspecto o tamaño del cuadro:** proporción entre el ancho y la altura de la imagen.

En el siguiente gráfico se representa una secuencia ficticia para mostrar las diferencias al digitalizar vídeo para preservación, es decir, reúne todos los fotogramas del archivo, y la que se realiza para la difusión, donde se eliminan algunos de los fotogramas.



Copia máster y soportes de distribución

Para finalizar se incluye a continuación un gráfico que representa el producto final de la digitalización, diferenciando la copia máster para la preservación de un documento cualquiera que sea éste, con una alta calidad; y por otro lado las copias de este máster para su distribución, cuya calidad es media/baja, ya que como se citaba al inicio de este Cuaderno prima la funcionalidad.



6. ALGUNAS RECOMENDACIONES FINALES

Recomendaciones de carácter general

- Como se señalaba al inicio del Cuaderno, no es uno de nuestros objetivos profundizar en el qué se debe digitalizar, sobre todo con el objetivo de la preservación. No obstante, sí nos parece interesante apuntar las recomendaciones del *Blue Ribbon Task Force*¹ al respecto, y que centra la mirada en cuatro escenarios diferentes que van desarrollando cada vez mayores activos digitales y que vale la pena preservar, y que consideran tienen un interés público de preservación a largo

¹ informe final del *Blue Ribbon Task Force* sobre la preservación y el acceso digital sostenible, denominado «[Sustainable Economics for a Digital Planet: Ensuring Long-term Access to Digital Information](#)» (Economía sostenible para un planeta digital: cómo garantizar el acceso a la información digital a largo plazo).

plazo: la información académica, los datos de investigación, el contenido cultural de propiedad comercial (como música o películas digitales) y el contenido Web producido en forma colectiva (como blogs).

- La mayoría de las entidades de referencia analizadas tiende a utilizar y recomendar formatos similares. Como regla general, para la preservación se utilizan en mayor medida los formatos no propietarios, reconocidos como estándares. Cuando se trata de difundir, y aunque existe una mayor flexibilidad, se utilizan en muchos casos formatos propietarios de uso común.
- Independientemente de nuestro objetivo en la digitalización, para la «copia máster» se tiende al uso de formatos que permitan mantener con la mayor fidelidad posible las propiedades del documento original, primando por encima de todo la calidad. En el caso de las copias de difusión o distribución, se tiende a buscar un mayor equilibrio entre la calidad y la funcionalidad.
- Como norma general, los formatos presentados como válidos para la preservación, también lo son para la difusión de los contenidos, si bien se aplica algún tipo de compresión o se reduce su resolución a fin de incrementar su funcionalidad.
- La principal divergencia encontrada en los formatos utilizados en la industria y en el ámbito de la preservación, reside en la utilización de medios de protección técnica.

A continuación, se presentan de forma esquemática las tendencias detectadas en los formatos de uso más común tanto para la preservación como para la difusión en los diferentes contenidos.

Recomendaciones para los contenidos en Texto

Para que la digitalización cumpla sus objetivos de preservación y sobre todo para mejorar la accesibilidad de los recursos, es necesario adaptar los contenidos a las nuevas tecnologías y dispositivos desarrollados desde el ámbito de la industria. En este sentido, aunque el PDF/A es el estándar más común para los documentos de texto con formato, cada vez son más las entidades que ofrecen contenidos en formatos de texto electrónico, siendo el EPUB el estándar de facto. Ambos formatos están basados en XML.

El **PDF/A** se presenta como el estándar aceptado para la creación de documentos digitales por su idoneidad para la presentación de contenidos que deban ser accesibles online y susceptibles de ser impresos, tanto aquellos basados en texto como los que incluyen imágenes u otros elementos (gráficos, diagramas...) que requieren de un diseño preciso. Su mayor inconveniente es que la anchura del texto no varía en función de la pantalla del dispositivo, es decir, si el documento no entra en la pantalla del dispositivo utilizado el usuario debe navegar horizontalmente lo que representa un obstáculo para su uso con los lectores de libros electrónicos (E-reader).

Por el contrario, el **EPUB** es el estándar de facto recomendado para el texto electrónico. Aunque puede soportar imágenes está más orientado a la publicación de texto, por ello no es el formato más adecuado para documentos que requieren un diseño preciso o están basados en imágenes.

Recomendaciones para los contenidos en Imagen

En la digitalización de imágenes, el TIFF sigue siendo el formato más utilizado en la preservación de documentos, aunque el JPEG 2000 va cobrando cada vez más adeptos por entidades de referencia a nivel internacional (por ejemplo, la Biblioteca del Congreso de los Estados Unidos) para la copia máster, siempre que la compresión sea sin pérdida.

En el cuadro siguiente, y a modo de referencia básica, se presentan las recomendaciones técnicas de la IFLA (*International Federation of Library Associations*) para la digitalización de imágenes; orientaciones que sirven de guía para la mayor parte de entidades de referencia tanto en la CAE como a nivel estatal o en la Unión Europea¹.

Recomendaciones IFLA para archivos de imagen	
Profundidad de Color	Nº de colores visibles
Preservación documentos impresos y manuscritos y otros materiales especiales (fotos, planos, grabados, mapas...)	Formato TIFF sin compresión / Resolución mínima 400 ppp / Color: Escala grises 256 gamas / 16,7 millones colores
Preservación prensa histórica	Formato TIFF sin compresión / Resolución mínima 400 ppp / Color: Escala grises (256 gamas)
Preservación prensa moderna	Formato TIFF sin compresión / Resolución mínima 300 ppp / Color: Escala grises (256 gamas)
Difusión	archivos con compresión JPEG

¹ Para aquellas personas que deseen profundizar en aspectos técnicos en digitalización de imágenes, existen documentos ([Biblioteca del Congreso de los EE.UU.: Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files \(Agosto 2010\)](#)) que presentan un mayor detalle sobre las especificaciones para cada tipología de documento o para cada parte del proceso: tratamiento de los originales, escaneado, OCR (ver [Kultura 2.0 – Píldoras de Formación \(abril 2011\)](#), OCR: [tecnología para el reconocimiento óptico de caracteres en una imagen](#).)

Recomendaciones para los contenidos en Audio

El WAV sigue siendo el formato de audio recomendado habitualmente para la preservación de archivos y el más ampliamente utilizado. No obstante, el formato BWF va incrementando su aceptación y uso, y es el formato recomendado actualmente por la IASA, la Biblioteca del Congreso de los Estados Unidos y la Biblioteca y Archivos de Canadá. El formato BWF está basado en el WAV y es compatible con cualquier software utilizado para reproducir este último. La diferencia estriba en que el BWF permite la incorporación de metadatos en el mismo archivo, y por ello puede resultar más adecuado si la finalidad es la preservación.

Recomendaciones IASA (International Association of Sound and Audiovisual Archives) para preservación	
•	Formato WAVE y BWF con fines archivísticos
•	Calidad recomendable: 96 kHz / 24 bit
•	Calidad mínima: 48 kHz /24 bit
•	Muchas entidades de preservación utilizan calidades superiores
•	Se aplican las mismas especificaciones para las grabaciones de voz

Recomendaciones para los contenidos en Vídeo

En el ámbito de la preservación de vídeos cuando no existen estándares normalizados o de uso común se siguen las tendencias de la industria, teniendo en cuenta su adecuación a los criterios de sostenibilidad. De esta forma, el uso creciente que tiene en la industria el formato JPEG 2000 encapsulado en un archivo MFX para las copias máster, ha llevado a su «adopción» por parte de entidades tractoras en la preservación como las ya citadas Biblioteca del Congreso de los Estados Unidos, la Biblioteca y Archivos de Canadá, o incluso más recientemente el INA—*Institut National de l'Audiovisuel*—de Francia.

La rapidez con la que evolucionan los formatos en el sector audiovisual, junto al incremento de la capacidad de almacenamiento de los archivos digitales y el ancho de banda, suponen un hándicap a la hora de definir unos estándares o establecer unas especificaciones técnicas que homologuen formatos. En todo caso, el DCI (*Digital Cinema Initiative*) y la BBC en el ámbito de la industria, y más recientemente el INA, la Biblioteca del Congreso de EE.UU. y la Biblioteca y Archivos de Canadá, han adoptado el formato MXF-JPEG 2000 como estándar para la preservación o para la copia máster de distribución en el caso de la industria.

Una de las principales ventajas del formato JPEG2000 es su capacidad para mostrar imágenes en diferentes resoluciones y tamaños, y para comprimir las imágenes sin pérdida. En el documento «[INA chooses JPEG2000 as its pivot format](#)», en el que se justifica la selección del JPEG2000 por parte del INA, se establece una tasa de bits de 100 Mb/s para definición estándar, y 200 Mb/s en alta definición, lo que supone un espacio de almacenamiento de 45 Gb y 100 Gb por hora de grabación respectivamente.

Para finalizar, se incluye un breve resumen de formatos según tipo de contenido y objetivo de digitalización.

Tipo de contenido	Preservación	Difusión
IMAGEN	TIFF (sin compresión) JPEG2000 (sin pérdida)	JPEG JPEG2000 (con pérdida) TIFF (con compresión)
TEXTO	EPUB (libro electrónico) PDF/A1 (ISO 19005-1)	EPUB PDF/A PDF HTML XHTML Open office Texto plano (ISO 8859-1)
AUDIO	BWF	MP3
VÍDEO	JPEG2000 (ISO/IEC 15444-4) (.mj2)	JPEG 2000 MPEG4 AVI MOV

7. RECURSOS COMPLEMENTARIOS

- **Conceptos básicos**
- **Instituciones internacionales de referencia en digitalización**
- **Selección de proyectos de interés y buenas prácticas**

CONCEPTOS BÁSICOS

Bit: es una síntesis de dos términos en inglés: *Binary digit*, que en español significan dígito binario, es decir, número (dígito) con dos posibles valores (binario). El término surge de usar las dos primeras letras de Binary con la última de digit: bit.

Compresión: se trata de la reducción de la cantidad de datos a transmitir o grabar, considerando que la capacidad de almacenamiento de los soportes es limitado. Hay dos tipos de compresión: Compresión sin pérdidas (se transmite toda la información eliminando la información repetida y agrupándola a fin de que ocupe un menor espacio); Compresión con pérdidas (se desprecia la información irrelevante, pudiendo en este caso perder calidad en el resultado final).

Formato digital: sistema de codificación de la información para su posterior almacenamiento o tratamiento en un soporte informático (documentos trabajados desde los programas del ordenador) o digital (captura de imágenes, por ejemplo, desde un escáner, una cámara de fotos digital).

Frecuencia de muestreo: número de muestras que se toman de la onda de sonido en un período de tiempo, se expresa en Khz (kilohercio).

Imagen vectorial: imagen digital formada por objetos geométricos independientes (segmentos, polígonos, arcos, etc.), cada uno de ellos definido por distintos atributos matemáticos de forma, de posición, de color, etc. Este formato es totalmente diferente a la imagen pixelada (rasterizada). Se puede ampliar el tamaño de una imagen vectorial como uno quiera sin que se perciba en ésta el efecto de escalado, que sí sufren las imágenes pixeladas.

Mapa de bits (o imagen rasterizada, imagen matricial o bitmap) una estructura o fichero de datos que representa una rejilla rectangular de píxeles o puntos de color, denominada raster, que se puede visualizar en un monitor, papel u otro dispositivo de representación.

Master: grabación original de la que se hacen las copias editadas.

Migración: método que consiste en la conversión de los documentos almacenados a nuevos formatos con el fin de no perder la información que contienen. Mediante este proceso se mantienen las características esenciales de los datos, aunque el cambio realizado entre tecnologías puede producir pequeños cambios que con el paso del tiempo y de consecutivas migraciones pueden verse acumulados.

Píxel: es la menor unidad homogénea en color que forma parte de una imagen digital, ya sea ésta una fotografía, un fotograma de vídeo o un gráfico (acrónimo del inglés *picture element*, «elemento de imagen»).

Profundidad de bits: número de bits utilizados para representar la onda sonora.

Profundidad de color: cantidad de bits de información que se necesitan para representar el color de un píxel en una imagen digital o en un framebuffer. Teniendo en cuenta que los octetos de 8 bits se aceptan universalmente como unidades básicas de información en los dispositivos de almacenamiento, los valores de profundidad de color son divisores o múltiplos de 8 (1,2,4,8,16,24 y 32), excepto la profundidad de color de 10 o 15, que se utiliza en determinados dispositivos gráficos.

Resolución: cantidad de puntos que componen una imagen, es decir, la capacidad de distinguir los detalles espaciales más finos. Se utiliza comúnmente en relación a imágenes de fotografía digital aunque también para describir el grado de nitidez (como antónimo de granular) de una imagen de fotografía convencional (o fotografía química). A mayor resolución la imagen presenta una mayor calidad visual (mps).

Streaming: hace referencia a la distribución de audio o vídeo por Internet. El usuario puede escuchar música o ver un vídeo en el momento que desee, sin necesidad de ser descargado previamente.

Sonido surround: hace referencia a la utilización de múltiples canales de audio encaminados a provocar efectos envolventes en la audiencia, ya sea proveniente de una película o de una banda sonora.

Tasa de bits (*bit rate*): velocidad de transferencia de los datos contenidos en el archivo audiovisual, se expresa en bits por segundo (bps). Determina la nitidez de la imagen ya que a mayor velocidad más información puede contener la imagen.

Velocidad de la imagen (*frame rate*): número de imágenes estáticas o fotogramas por segundo, se expresa en fps (*frames per second*). Existen distintos estándares para cine (24 fps) y los distintos sistemas de Televisión, etc.

Instituciones internacionales de referencia

Biblioteca del Congreso de los Estados Unidos

<http://www.digitalpreservation.gov/formats/index.shtml>

Biblioteca y Archivos de Canadá

<http://www.collectionscanada.gc.ca/digital-initiatives/012018-2210-e.html>

DCI - Digital Cinema Initiatives

<http://www.dcinovies.com/specification/index.html>

IASA - International Association of Sound and Audiovisual Archives

<http://www.iasa-web.org/audio-preservation-tc04>

[Directrices IFLA/UNESCO para el desarrollo del servicio de bibliotecas públicas](#)

INA- Institut National de l'Audiovisuel de Francia

[INA chooses JPEG2000 as its pivot format](#)

Selección de Proyectos de interés y buenas prácticas

[Biblioteca digital mundial](#). Entre sus objetivos está el promover el entendimiento internacional e intercultural; ampliar la cantidad y la variedad de contenidos culturales en Internet; facilitar recursos a los educadores, estudiosos y el público en general; permitir a las instituciones asociadas reducir la distancia digital dentro de y entre los países. Cuenta con el apoyo de la ONU.

[Europeana](#) permite explorar recursos y colecciones digitales de los museos, bibliotecas, archivos y archivos audiovisuales de Europa. Se pueden encontrar más de 15 millones de ítems así por ejemplo: imágenes, textos, sonidos y vídeos. Para permitir que los usuarios puedan realizar sus consultas en euskera, la Diputación Foral de Gipuzkoa ha desarrollado la interfaz de Europeana en euskera en el marco de la iniciativa Gipuzkoa 2.0

[Hispana](#) colecciones digitales de archivos, bibliotecas y museos conformes a la Iniciativa de Archivos Abiertos que promueve la Unión Europea. Cumple funciones análogas a las de Europeana en relación a los repositorios digitales, es decir, constituye un agregador de contenidos de las bases de datos de colecciones digitales. Da acceso a 3,024,912 objetos digitales de 151 repositorios. Dispone de un Directorio de colecciones digitales donde puede identificarse proyectos actualmente en ejecución.

[Biblioteca digital vasca](#): espacio de cooperación y colaboración interinstitucional que permite actualmente el acceso a las colecciones digitales de la Fundación Sancho el Sabio y la biblioteca del Parlamento Vasco. Señalar que las instituciones interesadas podrán depositar sus fondos digitales en el repositorio digital en el que se ha estado trabajando en los últimos meses, y el Departamento de Cultura se encargará de la custodia, mantenimiento, licencias, etc. necesarios para la correcta gestión de la documentación preservada. Dentro de la primera fase del proyecto se está trabajando con el Fondo Loyola y el Parlamento Vasco. La nueva [Biblioteca Digital de Euskadi](#) acopiará todas las colecciones de la Biblioteca digital vasca.

[Euskomedia Fundazioa](#) gestiona el mayor fondo digital existente sobre cultura vasca, de acceso libre y gratuito a través de su web. Destacan los más de 16.000 recursos digitales del fondo editorial, de los cuales la gran mayoría son a texto completo, los 150.000 artículos de la enciclopedia Auñamendi, o las más de 71.000 imágenes, 1.700 horas de audio y 144 horas de video y animaciones de la galería multimedia.

[Hedatuz](#), Biblioteca digital de ciencia y cultura vasca. Creado por Euskomedia Fundazioa, pone a disposición de la sociedad vasca todos los artículos de revista, monografías y obras especializadas editadas en el seno de Eusko Ikaskuntza desde el año 1918 y los números producidos por RIEV desde el año 1907. Son más de 5.000 artículos con acceso a texto completo y diversas temáticas: historia, antropología, artes, ciencias políticas, sociología, lingüística.

[Inguma](#). La UEU (Udako Euskal Unibertsitatea) presenta la Base de Datos de la Comunidad Científica Vasca que contiene en estos momentos 33.691 documentos (11.965 artículos; 4.344 prólogos, reseñas, artículos de libros o ediciones críticas; 1.850 libros; 387 tesis doctorales; 93 proyectos de investigación; 1.453 cursos, 8.231 conferencias y 4.698 asignaturas), adscritos a 8.076 autores. Por lo tanto, en estos momentos se pueden consultar en Inguma un total de 41.767 registros. Promovido por la Diputación Foral de Bizkaia y el Gobierno Vasco.

[La Filmoteca Vasca - Euskadiko Filmatagia](#) es un organismo que tiene como principal objeto, el archivo, conservación, restauración, y en su caso, exhibición de películas de cine en general pero, sobre todo, de las relacionadas con lo vasco. Asimismo, pretende reunir, custodiar y difundir la documentación, bibliografía y materiales cinematográficos que puedan ser de interés desde el punto de vista cultural o histórico.

[EITB](#). El portal EITB a la carta, que desde el 14 de junio ofrece en directo los contenidos de todas las emisoras de EITB, así como los programas emitido los últimos 7 días sobre los que ETB tiene el derecho de emisión, tiene previsto ir incluyendo en su oferta los contenidos de los más de 30 años de historia de la cadena recogidos en el archivo de EITB.