

Información entre bambalinas

Ángel Herráez

*Los metadatos son información acerca de la información.
A efectos prácticos para nosotros, se trata de datos almacenados de forma oculta aunque accesible y, lo más importante, están tipificados u organizados en campos normalizados que permiten su interpretación automatizada.
El autor nos explica por qué son tan importantes.*

*Metadata liberates us,
liberates knowledge*
David Weinberger¹

Cada vez más, los documentos que creamos y manejamos están en formato puramente digital y muchos de ellos los distribuimos abiertamente. Esto conduce a la proliferación e intercambio de documentos cuyo origen y autoría a menudo se pierden. Por otra parte, y más importante, también la gestión de toda la información de estos documentos se hace de modo automatizado: indexación, interconexiones entre documentos, análisis, extracción de palabras clave... En todos estos procesos, la presencia en el texto de una identificación del autor, que es suficiente para el lector humano, es difícilmente reconocible por los sistemas informáticos y, por tanto, se perderá tal información. Para colmo, no todos los profesores cuidan de incluir esa indicación de autoría en sus documentos docentes. En esta misma línea podemos sumar la declaración de las condiciones de uso (licencia), un tema tratado recientemente en esta sección.

Pretendo, pues, dedicar este espacio a compartir con vosotros algunas indicaciones sobre las posibilidades que tenemos para incluir este tipo de información en nuestros documentos digitales, de modo que esté disponible para el ojo entrenado y, en particular, para los sistemas automatizados de gestión documental —y aquí no hablo solo de software propio de bibliotecarios, sino del mero buscador de internet que indexará los documentos que publiquemos en la web.

La solución a estas cuestiones es lo que se ha denominado *metadatos*: información acerca de la información. A efectos prácticos para nosotros, se trata de datos que no se muestran directamente en el documento pero que están almacenados de forma oculta aunque accesible y, lo más importante, están tipificados u organizados en campos normalizados que permiten su interpretación automatizada. Por ejemplo: nombre de los autores y editores, entidad que publica, fecha de publicación, palabras clave, licencia de uso (fig. 1).

Cabe, pues, aprender dos cosas: cómo podemos leer esa información y cómo podemos generarla o modificarla. Hay, obviamente, una fuerte dependencia del

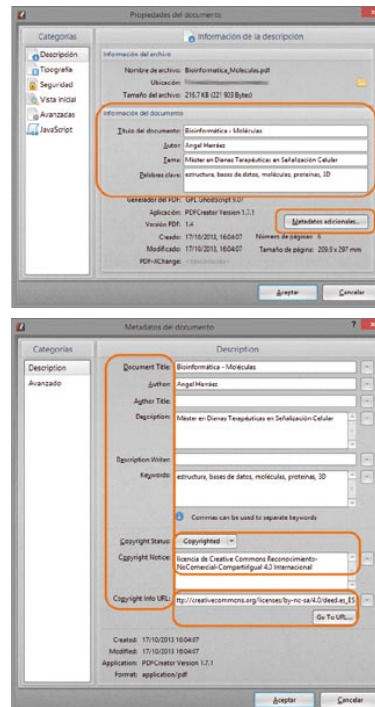


Figura 1. Ejemplo de metadatos disponibles en un documento (en formato pdf, editado con el programa PDF-XChange)¹³

Tabla 1. Acceso a la información de metadatos en diversos programas informáticos de edición de documentos

Programa	Acceso a los metadatos	Opciones configurables *
MS Office 2003	menú Archivo > Propiedades > pestañas «Resumen» y «Personalizar»	menú Herramientas > Opciones > pestaña «Guardar» > (Word) casilla «Preguntar por las propiedades del documento» (Powerpoint) casilla «Solicitar propiedades de archivo» (Excel) no disponible
MS Office 2007	menú principal > Preparar > Propiedades > se abre el «Panel de información»	
MS Office 2010-2013	menú Archivo > Información > Propiedades (panel a la derecha) Pueden elegirse «Mostrar panel del documento» y «Propiedades avanzadas»	
LibreOffice (v.4.2)	menú Archivo > Propiedades > pestañas «Descripción» y «Propiedades personalizadas»	menú Herramientas > Opciones > Cargar/Guardar > General > casilla «Editar propiedades del documento antes de guardarlo»
PDF Creator (pdfforge) ¹²	varios campos en el diálogo que se muestra antes de guardar	
Adobe Reader	menú Archivo > Propiedades (solo permite consultarlas, no modificarlas)	
PDF-XChange Editor ¹³	menú Archivo > Propiedades del documento > Descripción > «Información» y «Metadatos adicionales»	

* Si eres propenso a olvidarte de añadir tales detalles, puedes activar una opción para que se te pregunte siempre que guardas el documento.

software que empleemos para cada formato de documento, por lo que proporcionaré algunos ejemplos con aquellos a los que tengo acceso y son más habituales; podréis extrapolar la idea y los métodos a otros programas que utilicéis habitualmente.

► **Información susceptible de formar parte de los metadatos**

Aunque la lista podría extenderse en casos puntuales, reseñaremos brevemente los elementos fundamentales.

Título

La declaración de un título concreto, aparte de proporcionar una información básica, facilita, entre otras cosas, la cita.

Autoría

No solo es conveniente «firmar», o declarar el autor del documento, sino asimismo la

institución a la que pertenece –algo especialmente relevante para materiales educativos–. Dependiendo de los casos, se pueden añadir editores, revisores, editorial o entidad responsable de la publicación...

Fecha

En ocasiones juzgamos la vigencia de un documento dependiendo de cuándo se publicó, y este es uno de los datos con más frecuencia ausente en los documentos electrónicos, por falta de atención. Los metadatos pueden incluir fecha de publicación y, opcionalmente, alguna información suplementaria como una fecha de actualización o revisión.

Licencia de uso

Como se describía en el número de marzo,² es muy recomendable que nuestras obras especifiquen de forma clara y, a ser posible, normalizada una declaración de licencia. Esta puede incluirse

básicamente en una de tres categorías: *copyright* (todos los derechos reservados), Creative Commons o similar (algunos derechos reservados), dominio público (sin limitaciones).

Tema o asunto

Se puede incluir una referencia a la asignatura, curso, colección o cualquier entidad de organización a la que pertenezca el documento, o bien una descripción más detallada que el título.

► **Documentos en un procesador de textos, presentaciones, hojas de cálculo**

De forma muy directa es factible incorporar en estos formatos de documento los elementos de información indicados en el apartado anterior. Basta con localizar el apartado en el software que nos permi-

ta acceder a la información de metadatos (fig. 2). En la tabla 1 se recogen algunas indicaciones, a modo de receta abreviada, para encontrarlo en varios programas habituales.

Otra posibilidad adicional es usar una herramienta que simplifique la incorporación en el propio contenido (no en los metadatos) de la información normalizada de licencia; por ejemplo, existen algunos complementos que insertan la licencia Creative Commons que elijamos, con texto, icono e hipertexto.³

► Documentos en formato pdf

Personalmente encuentro preferible, salvo casos excepcionales, publicar siempre mis documentos en formato pdf, evitando los formatos específicos de procesador de texto, hoja de cálculo o presentación. Entiendo que la accesibilidad es mucho mejor con este formato, considerando que los usuarios emplean diversos sistemas operativos, pueden no tener los mismos programas que yo uso y además su apertura desde enlaces en páginas web siempre supone una demora, múltiples ventanas..., en general entorpece la navegación rápida y eficaz. Por otra parte, el formato pdf nos asegura que el aspecto del documento no se alterará al abrirlo en otros ordenadores, tabletas o incluso teléfonos, cosa no infrecuente con otros formatos y programas.

Aunque el programa conocido por todos es el visor de Adobe (cuyo nombre ha ido fluctuando entre Adobe Acrobat Reader y Adobe Reader), el formato PDF (*Portable Document Format*) es hoy en día un formato libre, normalizado por ISO. Esto añade un voto a favor de mi propuesta de emplear el pdf para todos los documentos, como un formato *universal y abierto*. Por otra parte, como consecuencia de tal normalización, los archivos pdf son accesibles usando software de otras compañías, de modo que existen diversos programas visores e incluso editores de archivos pdf, algunos son gratuitos, y algunos permiten modificar los metadatos de un documento creado previamente. Este último aspecto es especialmente

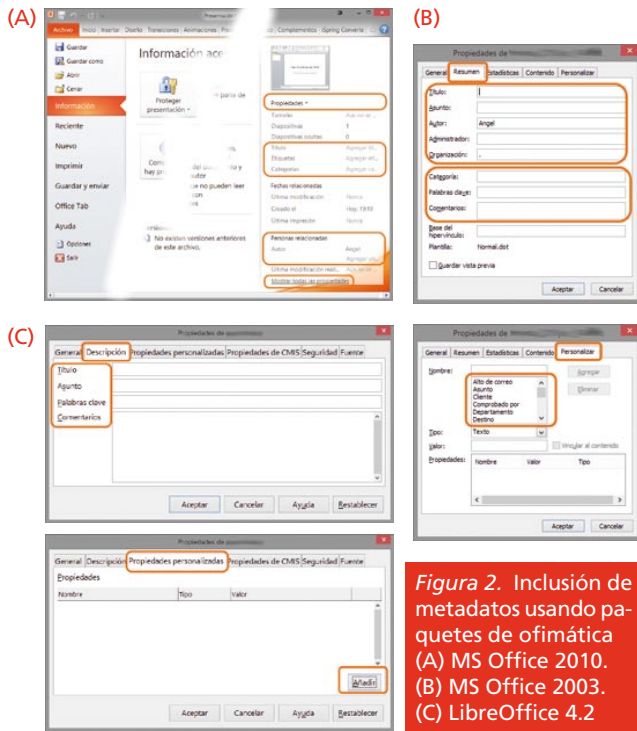


Figura 2. Inclusión de metadatos usando paquetes de ofimática (A) MS Office 2010. (B) MS Office 2003. (C) LibreOffice 4.2

tamente. Para atender a los programas que no ofrecen esa opción, se puede fácilmente instalar un “gestor de impresora virtual” que, en lugar de enviar a la impresora, envía a un archivo pdf que se guarda en nuestro ordenador; hay varios de ellos disponibles en internet, tanto gratuitos como de pago.⁴ En el proceso de grabar el archivo pdf puede haber una segunda oportunidad para añadir o rectificar los metadatos (fig. 3).

De todos modos, tras haber creado el documento pdf tenemos una tercera oportunidad, pues es posible editar sus metadatos para verificarlos, corregirlos o ampliarlos (véase un ejemplo en la fig. 1) y grabar de nuevo el archivo pdf con la información extra.

conveniente y acorde a los propósitos del artículo que nos ocupa.

En primer lugar, debe mencionarse que cuando se han introducido los metadatos en el documento original (por ejemplo, en el procesador de textos), lo habitual es que al generar la versión pdf se conserven esos metadatos. Esto depende, lógicamente, de qué herramienta emplees para producir los documentos pdf. Como sabréis, las versiones recientes de los paquetes ofimáticos más populares ya incluyen la opción de grabar en formato pdf direc-

► Páginas web

Desde sus comienzos, el formato HTML empleado para escribir las páginas web incluye la etiqueta META,⁵ destinada precisamente a contener metadatos, información sobre el documento que no se muestra en el navegador. Se trata de un elemento esencial para la indexación de las páginas web, aunque actualmente los buscadores de internet ya no solo se basan en esto, sino que «leen» todo el contenido de la página, incluso el de los documentos pdf.

Entre los campos de mayor utilidad en etiquetas meta podemos citar autor, idioma del documento, descripción, palabras clave, fecha de creación y duración estimada en que la información de la página no se modificará.

Podemos consultar esa información oculta—destinada en principio a navegadores, buscadores, robots web y otros sistemas automatizados de exploración de la red—utilizando alguna de las opciones de nuestro navegador. Por ejemplo, en Firefox un clic derecho sobre el fondo de página ofrece «Ver información de la página» (fig. 4) A falta de una ayuda como esta, podemos mostrar el código fuente desde el navegador y en él buscar las etiquetas meta.



Figura 3. Diálogo de producción de un archivo pdf usando una «impresora virtual», donde se pueden proporcionar metadatos (con el programa PDFCreator)¹²

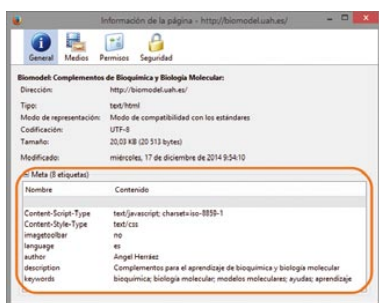


Figura 4. Consulta de metadatos de una página web (empleando el navegador Firefox)

La forma de añadir esta información depende del modo como escribas las páginas. Si utilizas un editor de páginas web, busca en el programa una opción del tipo «propiedades de la página» o bien «etiquetas html». Si conoces algo de código html, se introducen manualmente en el código fuente de la página (etiquetas <meta> dentro de la sección <head>). También puedes usar un asistente con una interfaz más amigable para que te genere el texto de las etiquetas y luego pegarlo en la página.⁶

Otro tipo de metadatos son aquellos asociados, no ya a la página web en su conjunto, sino a imágenes u otros elementos incluidos en aquella. Por ejemplo, a las imágenes insertadas en un documento html se les puede añadir un parámetro *alt*, que indica un texto alternativo, una descripción de lo que contiene la imagen; este es un elemento diseñado para la accesibilidad, para que las personas con limitaciones de visión puedan disponer de la información que transmite la imagen: el software de lectura de páginas, una herramienta básica para ellas, lee en voz alta el texto alternativo cuando alcanza una imagen. Otro elemento de metadatos es el parámetro *longdesc*, que dirige a una nueva página web destinada a proporcionar información detallada sobre dicha imagen; a diferencia de *alt*, *longdesc* también puede aplicarse a otro tipo de elementos multimedia. Un tercer contenedor de metadatos es el parámetro *title*, aplicable a cualquier elemento de la página, ya sea imagen, vídeo, sonido, una palabra, un botón... Su contenido se muestra en el navegador como una viñeta emergente (*tooltip*) cuando detenemos el puntero sobre el elemento en cuestión. Además de la función primaria de estas tres utilidades para orientar al lector de la página, los buscadores también leen estos campos como parte de la página.

Para usar estos parámetros de las etiquetas html correctamente es preciso un conocimiento del código fuente html o bien localizar estas funciones (una vez que sabemos que existen y lo que significan) en nuestro programa editor de páginas.

► **Imágenes**

La primera solución es la inclusión de la información como parte de la imagen, típicamente con un pequeño icono o logotipo y algún texto, en una esquina o como marca de agua. Quizás no podemos considerar que esto sean metadatos en sentido estricto, en cuanto no van a ser reconocidos por ningún sistema que no sea el cerebro humano, pero sí cumplen el propósito de comunicar esa información a los usuarios. Una objeción notable es que puede alterarse la estética o la *limpieza* del material gráfico.

«Es posible también incorporar metadatos a determinados archivos de imagen.»

Con respecto a las imágenes incorporadas dentro de una página web, en la sección anterior ya se han comentado varias soluciones, relacionadas con el uso del texto alternativo o adicional a la imagen.

Finalmente, dependiendo del formato del archivo de imagen y del software que empleemos para editarla, existen posibilidades de incorporar verdaderos metadatos dentro del archivo. La cuestión es algo técnica y muy vinculada al programa de edición específico, así que plantearé una descripción general como orientación.

El formato *jpeg* (o *jpg*) admite la inclusión de información adicional a la que constituye la imagen; aunque no te hayas detenido a pensarlo, probablemente tengas experiencia de esto: las fotos hechas con cámara digital (ah, pero ¿las hay de otro tipo?) incluyen al menos la fecha y hora en que se tomaron, el modelo de la cámara y algunos de sus parámetros ópticos como distancia focal, sensibilidad de la «película»... Si haces la foto con un teléfono o una

tableta, quizás incluso tu localización GPS cuando tomaste la foto. Todo esto se hizo posible mediante una especificación conocida como datos EXIF. Los programas de retoque de fotos son capaces de leerla, e incluso el propio administrador de archivos de tu ordenador (prueba en «propiedades del archivo»). Aprovechando este principio, algunos programas pueden incluir información en ese apartado del archivo.

El formato *png* permite igualmente la inclusión de información de metadatos.⁷ A tal extremo llega esta opción que, por ejemplo, el programa Jmol para la visualización de estructuras moleculares⁸ es capaz de guardar archivos en formato *png* que contienen, además de la captura de imagen, los datos de coordenadas atómicas de la molécula, su estilo de representación... la *escena* completa del modelo molecular. La mayoría de programas lo ven como un simple archivo de imagen, pero Jmol puede leer tal archivo y regenerar el modelo 3D de la molécula tal como lo teníamos al grabar el archivo *png*. El secreto es que en esa zona reservada a datos se ha incluido toda la información del modelo molecular. Podríamos decir que, más que metadatos, ¿esto ya son ultradatos! En una aproximación más convencional, la idea es que podamos incorporar la información de autoría, licencia, descripción, etc., dentro del archivo de imagen *png*. Para ello habrá que utilizar algún editor de imágenes con esta prestación.⁹ De todos modos, parece que aún es una característica con una compatibilidad imperfecta.

El formato *svg* es oficialmente un formato *de futuro*; en Wikimedia Commons y Wikipedia es el formato recomendado para las imágenes; posee ventajas por ser un formato abierto y vectorial –por tanto, escalable sin pérdida de calidad–. A pesar de todo ello, su implantación no es aún

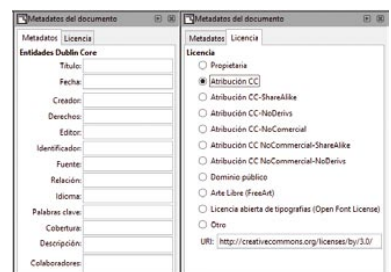


Figura 5. Inclusión de metadatos en un archivo de imagen en formato *svg* (editado con *Inkscape*)¹⁰

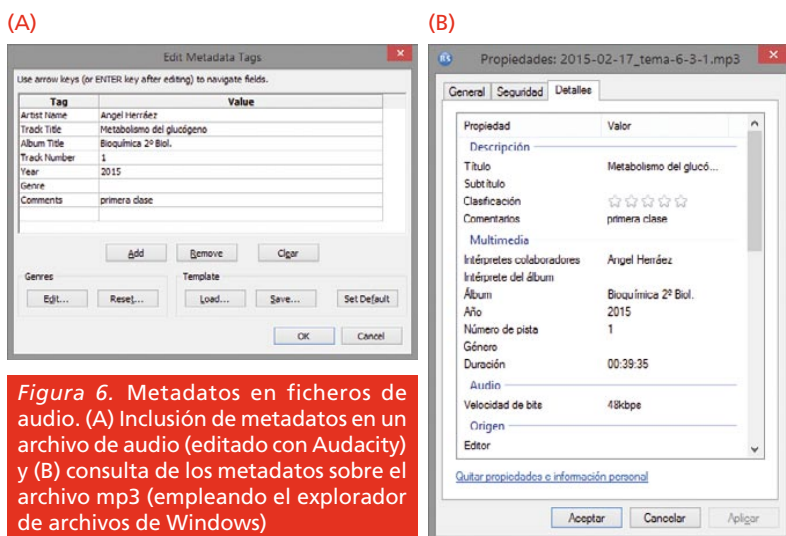


Figura 6. Metadatos en ficheros de audio. (A) Inclusión de metadatos en un archivo de audio (editado con Audacity) y (B) consulta de los metadatos sobre el archivo mp3 (empleando el explorador de archivos de Windows)

extensiva. Siendo un formato de texto, estructurado en campos y etiquetas, y por ello muy flexible, permite perfectamente la inclusión de todo tipo de información. La figura 5 muestra un ejemplo de las opciones disponibles empleando un software de referencia para este formato.¹⁰

► **Vídeos, animaciones y sonido**

En estos casos se hace probablemente más difícil incorporar y verificar la información de metadatos. Dependerá fuertemente del software empleado para la edición del vídeo o audio. También el formato final del archivo puede ser un condicionante, pues podría admitir o no esa información extra. No puedo entrar en detalles, que no domino y en cualquier caso serían especializados. Quienes estéis habituados a preparar este tipo de materiales podréis investigar su documentación y opciones en busca de la forma de añadir los metadatos (fig. 6).

Otro problema añadido es la visibilidad de los metadatos para los usuarios: aunque el archivo de vídeo o audio los lleve incluidos, es más improbable que el reproductor sea capaz de mostrarlos. Una solución adicional, quizás algo casera, especialmente en cuanto nos interesa hacer patente la autoría y la licencia, es incluir dicha información en el propio vídeo o la grabación de sonido. Bueno, me desdigo, no va a resultar tan casera: ¡los vídeos comerciales suelen incluir los títulos de crédito tanto al principio como al final de la película! Aplicando esta idea, al editar el vídeo podemos añadir una

breve pantalla (o mejor, dejarla fija al término del vídeo) con la identidad de los autores y la licencia de uso.¹¹

Como opción adicional, algunos sitios de internet dedicados al almacenamiento y compartición de recursos audiovisuales poseen un sistema propio de etiquetado de los materiales, incluyendo entre ellos la declaración de licencia. Esto quiere decir que aunque tu vídeo o imagen no lleven incluida en su interior la información —o la tenga pero invisible al usuario—, al acceder a él en la web del proveedor sí podrá verse la información que tú hayas añadido al subirlo y completar dicho etiquetado. En el caso de las licencias, es frecuente que se ofrezca una lista de ellas, por lo que no necesitas escribir toda la terminología, icono, enlace, etc., sino simplemente elegir la que prefieres. Además, como una consecuencia más de ese sistema de clasificación, en estos casos suele ser posible para los usuarios buscar materiales en ese repositorio web de acuerdo con el tipo de licencia.

Termino con el deseo de que esta exposición os haya resultado interesante, inspiradora y amena, y el reto para todos de que entre corrección y corrección de exámenes encontréis un huequecillo para revisar vuestros materiales y añadirles ese «tesoro escondido» de autoría, licencia de uso... ¡metadatos! #

.....
Ángel Herráez
 BIOQUÍMICA Y BIOLOGÍA MOLECULAR,
 DEP. DE BIOLOGÍA DE SISTEMAS,
 UNIVERSIDAD DE ALCALÁ

► **Bibliografía y notas**

- 1 David Weinberger (2008) *Knowledge at the End of the Information Age*. Bertha Bassam lecture, University of Toronto, 7 February 2008. Disponible en <http://bit.ly/1JMwFhK> (Consultado 5 mayo 2015)
- 2 Ángel Herráez (2015) *Mío, suyo... ¿nuestro?* Revista SEBBM 183, 34-38. Disponible en <http://www.sebbm.com/revista/articulo.asp?id=10818&catgrupo=270>
- 3 (a) *Creative Commons Add-in for Microsoft Office*. <http://bit.ly/1JjKqgh> Válido para Word, PowerPoint y Excel, versiones 2007 a 2013 para Windows (solo en inglés)
 (b) *LibreOffice plugin: paste images with credit*. <http://bit.ly/1QOv3bz> Al pegar en un documento una imagen que procede de internet y tiene metadatos, añade una línea de créditos con la información.
- 4 Para producir archivos pdf desde cualquier programa instalado, puedo recomendaros *PDF Creator* ofrecido en SourceForge (<http://sourceforge.net/projects/pdfcreator/>) y en la sede web del productor, pdfforge (<http://pdfforge.org/>). Es gratuito, sin publicidad y ofrece excelentes prestaciones. Ofrezco disculpas a quienes no uséis Windows, pues desconozco si este software está disponible para otros sistemas; claro que probablemente no sea necesario en vuestros sistemas.
- 5 *Meta element* (1 mayo 2015) En Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/w/index.php?title=Meta_element (Consultado 5 mayo 2015).
- 6 Miarroba Networks S.L. (2015) *Generador de Meta Tags*. <http://metatags.miarroba.es/> (Consultado 5 mayo 2015).
- 7 Exiv2 Community (2015) *The Metadata in PNG files*. <http://bit.ly/1QON89m> (Consultado 5 mayo 2015).
- 8 *Jmol: un visor Java de código abierto para estructuras químicas en tres dimensiones*. <http://jmol.org> y <http://wiki.jmol.org/>
- 9 Algunas soluciones son
 (a) *ImageMagick* (multiplataforma) <http://imagemagick.org/>
 (b) *TweakPNG* (Windows) <http://entropymine.com/jason/tweakpng/>
 (c) *PNGCommentator* (MacOS) <http://echomist.co.uk/software/PNGCommentator.html>
- 10 *Inkscape* <http://inkscape.org/>
- 11 (a) Puede verse un pequeño ejemplo al principio y al final de la animación en <http://bit.ly/sintasaAG>
 (b) La wiki de Creative Commons proporciona algunas orientaciones para incluir licencias en vídeos y otros medios: https://wiki.creativecommons.org/Marking_your_work_with_a_CC_license
- 12 pdfforge GmbH (2015) *PDFCreator - the free PDF converter tool*. <http://pdfforge.org/> pdfcreator (Hay otros muchos programas equivalentes, que permiten grabar un documento pdf a partir de cualquier programa utilizando la opción de imprimir)
- 13 Tracker Software Products (2015) *PDF-XChange Editor*. <http://tracker-software.com/product/pdf-xchange-editor>