CREACIÓN DE UNIDADES DE ANÁLISIS FORENSE EN BIBLIOTECAS

Theo Wilderbeek y Miquel Térmens

Theo Wilderbeek es doctorando del *Departamento de Biblioteconomía y Documentación* de la *Universidad de Barcelona* (*UB*), graduado en información y documentación por la *Universitat Oberta de Catalunya* (*UOC*) y máster en gestión de contenidos digitales por la *UB*. Realiza su tesis doctoral sobre técnicas de análisis forense aplicadas a bibliotecas.
http://orcid.org/0000-0002-8378-1169

theo.wilderbeek@gmail.com

Miquel Térmens, doctor en documentación, es profesor del *Departamento de Biblioteconomía y Documentación* de la *Universidad de Barcelona* y director científico del *Centro de Digitalización* de la *Universidad de Barcelona*. Es especialista en digitalización y en preservación digital de documentos.
http://orcid.org/0000-0002-7305-3424

termens@ub.edu

Universidad de Barcelona, Departamento de Biblioteconomía y Documentación
Melcior de Palau, 140. 08014 Barcelona, España

Resumen

Las técnicas de análisis forense digital, de aplicación en investigación criminal, también se pueden usar en las bibliotecas para acceder a información digital almacenada en soportes o formatos obsoletos. Se analizan distintos ejemplos de departamentos de análisis forense creados por bibliotecas y se describen los elementos de hardware y software mínimos con los que se podría montar una unidad de análisis forense en cualquier biblioteca. Con este fin se presentan dos posibles configuraciones de equipamiento y se dan recomendaciones sobre organización del flujo de trabajo para la recuperación de antiguos discos duros y disquetes.

Palabras clave

Preservación digital; Análisis forense digital; Archivos personales; Bibliotecas; Hardware; Software.

Title: Creating forensic units in libraries

Abstract

Forensic analysis techniques, usually applied in criminal research, could also be used in libraries to access digital information stored in obsolete formats or storage devices. This article analyses some examples of forensic research departments created by libraries, and describes the minimal hardware and software elements required to set up a library unit specialized in forensic analysis. Two possible equipment settings are introduced and recommendations are given on how to organize a workflow to recover information stored in floppy disks, diskettes and old hard drives.

Keywords

Digital preservation; Digital forensics; Personal archives; Libraries; Hardware; Software.

Wilderbeek, Theo; **Térmens, Miquel** (2015). "Creación de unidades de análisis forense en bibliotecas". *El profesional de la información*, v. 24, n. 1, enero-febrero, pp. 44-54.

http://dx.doi.org/10.3145/epi.2015.ene.06

1. Introducción: el análisis forense

El análisis forense digital fue definido en 2001 como "el uso de métodos demostrados y derivados hacia la preservación, colección, validación, identificación, análisis, interpretación, documentación y presentación de evidencias digitales derivadas de fuentes digitales con el propósito de facilitar o fomentar la reconstrucción de acontecimientos que hayan

Artículo recibido el 06-06-2014
Aceptación definitiva: 08-10-2014

resultado ser criminales, o ayudar a prever acciones no autorizadas que se hayan demostrado como perjudiciales para las operaciones previstas" (**Palmer**, 2001). De forma más simple, se da el nombre de análisis forense digital al conjunto de técnicas y procedimientos que permiten acceder y analizar datos digitales que son prueba de actos delictivos. Son por tanto métodos propios de la investigación criminalística, que es donde nacieron y donde se han aplicado hasta el momento. Se trata de la traslación al mundo digital de los principios de la investigación forense tradicional: encontrar pruebas científicas de un delito cometido. El concepto de análisis forense digital es ampliamente conocido dentro de la investigación criminal dedicada a esclarecer delitos informáticos, siendo usado tanto por empresas privadas de investigación como por las fuerzas policiales.

> "Es posible montar una unidad de análisis forense digital de bajo o medio coste en instituciones medianas o grandes y se pueden conseguir grandes ventajas disponiendo de ella"

En el ámbito español, entre las fuerzas policiales y de seguridad del estado que usan estas metodologías destacan la *Brigada de Investigación Tecnológica* del *Cuerpo Nacional de Policía*, el *Grupo de Delitos Telemáticos* de la *Guardia Civil*, la *Sección Central de Delitos en Tecnologías de la Información* de la *Ertzaintza*, la *Unidad Central de Delitos Informáticos* de los *Mossos d'Esquadra*, y el *Centro Criptológico Nacional*.

Casi todo el mundo guarda sus archivos personales en formato digital (ya sean documentos, fotografías, vídeos o datos en bruto) y la tendencia es la sustitución del papel por los soportes informáticos. Es lógico suponer que dentro de poco tiempo el uso del papel será minoritario en los archivos personales, por lo que las bibliotecas deberán estar preparadas para los desafíos que plantea la preservación de contenidos digitales.

Las técnicas y métodos que se usan en el análisis forense digital representan una oportunidad interesante para las bibliotecas y los archivos patrimoniales, ya que dentro de las colecciones personales que se reciben procedentes de donaciones se está incrementando la presencia de soportes informáticos como disquetes, discos duros, cd-roms e incluso de ordenadores completos (**John**, 2012). De la misma manera que gracias al análisis forense un agente policial puede recuperar las pruebas de un crimen sin alterar los datos originales, los bibliotecarios pueden beneficiarse del análisis forense cuando se enfrentan al problema de acceder a contenidos nacidos digitales. Estos contenidos están guardados en dispositivos que plantean desafíos como:

- hardware: soportes antiguos como los disquetes no se pueden leer en los ordenadores actuales;
- software: los datos podrían haber sido creados con herramientas incompatibles con los sistemas actuales;
- el sistema operativo utilizado originalmente: en algunas ocasiones la visualización del contenido podría requerir de soluciones alternativas como la emulación o el uso de ordenadores antiguos que faciliten el acceso a los datos.

El análisis forense digital que permite a la policía acceder a los datos informáticos de un delito, puede tener una aplicación al facilitar que las bibliotecas y los archivos puedan acceder a datos informáticos obsoletos con el fin de conservarlos.

Este artículo expone las soluciones de hardware y software que ofrecen las tecnologías de análisis forense digital y su forma de integración en bibliotecas. Mediante el análisis de experiencias llevadas a cabo en el ámbito anglosajón (Estados Unidos, Reino Unido y Australia), se plantean dos posibles implementaciones: una a nivel básico para bibliotecas con un presupuesto reducido, y otra a nivel avanzado para centros que pueden permitirse realizar una inversión más elevada.

2. La transformación de los fondos personales

Históricamente, diferentes personalidades han donado sus fondos personales a bibliotecas, especialmente a organismos encargados de la preservación del patrimonio bibliográfico como bibliotecas nacionales, y también bibliotecas universitarias. La tendencia actual es claramente la paulatina desaparición del papel y su sustitución por el contenido nacido digital. Este cambio lleva asociado el problema de una adecuada gestión de los ficheros recibidos en formatos obsoletos o actuales (*Paradigm project*, 2007; **Kirschenbaum**; **Ovenden**; **Redwine**, 2010). Si a esto le sumamos el uso de soportes de almacenamiento obsoletos, el acceso a los datos originales es muy complejo, existiendo además un riesgo de pérdida de los mismos por el tiempo transcurrido desde que se guardaron.

Imaginemos el caso de una tesis doctoral que fue escrita con ordenador en 1989. El fichero de texto fue creado con *WordStar*, un software de procesamiento de textos que funcionaba bajo el sistema operativo *MS-Dos* y no disponible desde hace más de 20 años. Probablemente el fichero se almacenó en un disquete de 5,25 pulgadas; un soporte imposible de leer con los ordenadores actuales, pues ni se fabrican unidades de lectura compatibles, ni las placas base de los ordenadores están preparadas para conectarlas e interpretarlas. Aunque el disquete esté conservado en la biblioteca de la universidad donde el doctor defendió su tesis, los problemas tecnológicos citados impedirán que se pueda acceder a su contenido, sin olvidar que además ese disquete tiene una alta probabilidad de contener errores de lectura debido a su fragilidad material y a la degradación provocada por el paso del tiempo. Tan sólo en el caso de contar con una unidad forense digital la biblioteca estaría en condiciones de capturar los contenidos originales.

Siguiendo con este ejemplo se podrá decir que los problemas descritos se podrían haber evitado si la biblioteca en su momento hubiera realizado una adecuada vigilancia tecnológica y hubiera migrado los contenidos del disquete antes de que su acceso resultara problemático. Una migración del soporte (por ejemplo pasando de disquete a disco duro) y del formato del fichero (por ejemplo pasando de *WordStar* a pdf) hubieran sido las prácticas más adecuadas. Desgra-

ciadamente este procedimiento no es aplicable en el caso de la recepción de archivos patrimoniales donados por sus autores o sus herederos.

Una variante de la recuperación de archivos personales es la necesidad de acceder a cualquier tipo de dato informático contenido en soportes o formatos obsoletos, conocida como arqueología digital. No son pocas las bibliotecas y los archivos que en sus fondos almacenan cintas magnéticas o disquetes magnéticos que años atrás fueron depositados por instituciones públicas o por empresas y a los que ahora no es posible acceder. En muchos casos el problema empieza con la identificación del tipo de soportes, datos y programas que se encuentran presentes, para iniciar luego una investigación que permita acceder a estos datos del pasado. Se trata de un tipo de investigación que presenta claros paralelismos con la arqueología: se dispone de evidencias del pasado, pero no se sabe si estas son completas o no, no se sabe cuál puede ser su significado y al manipularlas existe un gran riesgo de destruirlas para siempre, sin que lleguen a comunicar nada concreto.

El análisis forense digital aplicado a las bibliotecas no se contenta con la recuperación de los ficheros, pues en ocasiones, cuando se trabaja sobre discos internos de ordenador, se puede llegar a perseguir el objetivo de recrear todo el funcionamiento de un ordenador de una persona determinada. Así, si conseguir acceder a los ficheros de un escritor se puede considerar que es el equivalente a acceder a los archivos en papel de escritores de otras épocas, dando un paso más, recrear el funcionamiento de su ordenador –con su estructura de directorios, su correo, su papelera, sus fotos almacenadas, etc.– es conocer de forma directa como trabajaba en su momento, de forma parecida a conocer como un autor pretérito tenía dispuesta su mesa, sus anaqueles, donde guardaba el tintero, etc.

> Las unidades de análisis forense podrían ser soportadas por más de una institución para resolver sus respectivas necesidades

Pero el análisis forense ya no se circunscribe al pasado, también quiere aportar soluciones para la gestión de los datos digitales del futuro. En un mundo en el que cada vez se genera más información, su gestión es cada vez más difícil, en especial si se espera realizarla con participación humana. Las técnicas de ingesta, acceso y almacenamiento forense de discos enteros pueden ayudar a facilitar la gestión de grandes volúmenes de datos. Un ejemplo claro y que se avecina es la gestión de los datos en bruto procedentes de proyectos de investigación y que los planes de gestión de datos (*data management plan*) obligaran a preservar para su reutilización futura.

La aplicación de las técnicas forenses requiere disponer de equipamiento específico y de personal con la adecuada formación. Una parte de esta puede venir con la incorporación de personal informático, pero también será imprescindible que profesionales bibliotecarios se formen en estos ámbitos.

Figura 1. Ejemplos de disquetes obsoletos: 5,25", 3,5" y Zip

3. Ejemplos de unidades forenses en bibliotecas

En los últimos años algunas grandes bibliotecas han empezado a usar las técnicas de análisis forense y para ello han creado departamentos especializados. Presentamos a continuación algunos casos.

3.1. *Stanford University Libraries*

Crearon en 2009 su propia unidad forense debido al gran volumen de ítems nacidos digitales que almacenaban –más de 18.000– y que requerían ser tratados (**Kirschenbaum**; **Ovenden**; **Redwine**, 2010; **Olson**, 2011). Disponen de dos unidades *FRED* (*Forensic recovery of evidence device*) de la empresa *Digital Intelligence*, licencias de software forense comercial (*Encase*, de *Guidance Software* y *FTK*, de *AccessData*) y una cámara réflex digital para documentar el estado inicial de los soportes. En el mismo año, las *Stanford University Libraries* se sumaron al proyecto colaborativo *AIMS*, que tiene como objetivo la definición de directrices de buenas prácticas para la gestión del material nacido digital (*AIMS Work Group*, 2012).

Los primeros esfuerzos de preservación de contenidos se centraron en el archivo del paleontólogo Stephen Jay Gould, cuyos archivos personales contenían 60 disquetes de 5,25 y de 3,5 pulgadas, tarjetas perforadas y tres cintas magnéticas. Tras diversas pruebas, se crearon imágenes de disco de los disquetes mediante el software gratuito *FTK Imager*, cuyas funciones incluyen la confirmación de que se ha creado correctamente una imagen de disco y la generación de un listado de los contenidos del soporte. (**Edwards**; **Chan**; **Olson,** 2010)

En lo que se refiere al tratamiento de los datos, el software *FTK* se encargó de la extracción de metadatos técnicos como el tamaño de los ficheros, fechas de creación, formato de fichero, etc. En este proceso se identificaron también datos sensibles como los correspondientes a tarjetas de crédito o el número de la seguridad social, que fueron marcados con la finalidad de bloquear su consulta por el público general y así preservar la confidencialidad de estos datos privados. Para identificar los contenidos de ficheros con formatos obsoletos (que no pueden ser abiertos con el software actual) se utilizó el visualizador interno del *FTK* que permite la lectura de más de doscientos formatos. Finalmente, se generaron informes en formato xml/html de los ficheros y

se exportaron al repositorio *Hypatia*, el cual permite el acceso a contenidos nacidos digitales de diversos donantes, además de los de Stephen Jay Gould. El repositorio permite el acceso a una parte de los ficheros que se han preservado, a una fotografía del soporte original y a la imagen de disco o a los ficheros que se encontraban en el soporte.
http://hypatia-demo.stanford.edu

3.2. *Bodleian Library*

Forma parte de la red de bibliotecas de la *University of Oxford* (Reino Unido), cuyas colecciones suman más de 11 millones de documentos. La biblioteca de investigación principal es la *Bodleian Library*, pero sus funciones son más amplias, ya que se encuentra incluida en la *Agency for the Legal Deposit Libraries* que tiene la misión de recibir el depósito legal del Reino Unido junto con la *British Library*. Como es lógico, la *Bodleian* tiene un fondo importante de material nacido digital y para asegurar su conservación y acceso se creó una sección en la biblioteca con el objetivo de establecer un repositorio de preservación digital, el *BEAM* (*Bodleian Electronic Archives and Manuscripts*).

Uno de los casos con los que se ha enfrentado es el del archivo de Barbara Castle, política del *Partido Laborista* británico. Entre sus materiales se encontraban 31 disquetes de 3 pulgadas compatibles con el ordenador *Amstrad PCW*, de 1985 (**Laing**, 2004). El contenido de los disquetes se consiguió recuperar gracias a una migración de los datos de texto escritos originalmente con el procesador de textos *LocoScript*, de *Locomative Software* (**Thomas**, 2011). La migración se consiguió gracias a la adquisición de un ordenador *Amstrad PCW*, que se conectó a un ordenador portátil que ejecutaba una máquina virtual con un sistema operativo *Windows 95*, mientras que un programa especial transformaba los datos a formato ascii o bien a rtf (*rich text format*).

Generalmente, el flujo de trabajo de preservación de contenidos incluye la realización de una fotografía del soporte y la creación de una imagen de disco, en la que se verifica que los datos de la imagen son una representación fidedigna del ítem original mediante un valor *checksum*[1]. Las imágenes de disco se almacenan en el repositorio de preservación junto con los metadatos correspondientes y un listado de ficheros que contiene. Por el momento, el repositorio *BEAM* no está abierto a la consulta, pero la *Bodleain Library* prevé que una parte del material estará disponible a los investigadores en un futuro próximo.

3.3. *National Library of Australia* (*NLA*)

En 2008 inició un proyecto de preservación, el *Digital preservation workflow project*, debido al importante crecimiento de material nacido digital en sus colecciones. El resultado fue la aplicación *Prometheus*, consistente en "un proceso escalable y semiautomatizado para transferir datos desde soportes físicos a un sistema de almacenaje de preservación digital" (**Del-Pozo**; **Elford**; **Pearson**, 2009). El hardware consiste en una unidad móvil que se puede conectar a cualquier estación de trabajo del personal de biblioteca mediante una conexión SATA. El software es libre y gratuito y permite tanto la creación de imagen de disco (mediante los programas *dd* y *cdrdao*), la verificación de la imagen mediante el valor *checksum* (programa *Jacksum*), la identificación de ficheros (programa *Droid*), la validación de ficheros (programa *Jhove*) y la extracción de metadatos (programa *NLNZ Metadata extractor*).

El flujo de trabajo con la aplicación *Prometheus* incluye la ingesta de las imágenes de disco y de sus metadatos en el repositorio de uso interno, el cual almacena dos copias del contenido: una es la imagen de disco y la otra los ficheros extraídos del soporte. Los contenidos se incorporan al almacén digital *DOSS* (*Digital object storage system*), donde se almacenan la mayor parte de las colecciones digitales (**Verheul**, 2006). El sistema actualmente no puede procesar ciertos soportes, como discos duros externos USB, debido a límites de espacio de *Prometheus*, pero en cambio el equipo de preservación sí trabaja con medios poco comunes, como discos Zip, tarjetas de memoria flash o disquetes de 5,25 pulgadas.

3.4. *Yale University*

La sección de la *Yale University* (Estados Unidos) con un mayor número de contenidos digitales es la de *Manuscripts and Archives*, fundada en 1969. Sus colecciones incluyen documentación sobre historia, arquitectura, ciencia, medicina y cultura. Esta universidad forma parte del proyecto colaborativo *AIMS*, por lo que algunos datos de sus colecciones se encuentran en el repositorio *Hypatia*, como por ejemplo el archivo personal de James Tobin, premio Nobel y profesor de economía, cuyos fondos incluyen 25 disquetes de 3,5 pulgadas. Las referencias a los disquetes se guardaron en un fichero xml codificado bajo el estándar *EAD* (*encoded archival description*) que se subió al repositorio *Hypatia* (*AIMS Work Group*, 2012).

> "El trabajo con los materiales digitales obsoletos debe empezar desde el momento de la donación o adquisición, con el fin de asegurar sus condiciones legales y obtener toda la información posible sobre el entorno en el que fueron creados"

En el flujo de trabajo se utiliza hardware especializado en preservación digital, como la tarjeta controladora *Kryoflux* y dispositivos *write-blocker* conectados a las estaciones de trabajo, con los que se crean las imágenes de disco, se generan los valores *checksum*, se extraen los metadatos (en este caso con el programa *Fiwalk*) y se procede a la ingesta de los contenidos (imagen de disco y metadatos) en el repositorio *Rescue* de uso interno. El acceso a los materiales está abierto a los investigadores.

3.5. *Emory University*

En 2010 la *Emory University* (Estados Unidos) recibió una donación muy especial: el archivo personal del escritor Salman Rushdie, que incluía un ordenador *Apple Macintosh* de sobremesa, dos ordenadores portátiles *Macintosh* y un disco duro externo. El material nacido digital sumaba 40.000 ficheros y 18 gigabytes de datos (**Loftus**, 2010). La adquisición

de estos aparatos motivó la creación de un grupo de trabajo específico para la preservación de materiales nacidos digitales, el *Born-Digital Archives Working Group* (**Carroll** *et al.*, 2011), cuya misión consistió en encontrar la mejor forma de presentar y preservar la información respetando la privacidad del autor.

El personal asignado al grupo acordó finalmente el tratamiento del material como colección híbrida, formada por materiales físicos y digitales. La estrategia elegida de preservación fue la emulación de la interfaz original a través del software de código abierto *SheepShaver*, que permite consultar la colección, manteniendo el *look-and-feel* del sistema original *Mac*, en las instalaciones de la universidad. El motivo principal de esta solución fue minimizar los riesgos de incurrir en errores, pérdida y autenticidad de los datos. Por otro lado no se pierde la posibilidad de una futura migración porque en todo caso se conservan los datos originales.

Una clara muestra del éxito de esta iniciativa es que para facilitar la investigación de materiales nacidos digitales la institución está creando una unidad de análisis dedicada a la realización de imágenes de disco, a la recuperación de datos dañados, a la asignación de valores *checksum* y al mantenimiento de una colección de materiales nacidos digitales.

4. Requerimientos

En el apartado anterior hemos visto algunos ejemplos de instituciones que ya están aplicando el análisis forense dentro de sus procesos de gestión documental. También hemos visto que se han realizado grandes inversiones para dotarse de los equipos y el software más avanzado. Ello puede llevar a pensar que este tipo de metodologías sólo están al alcance de unas pocas instituciones (**Erway**, 2012; **Barrera-Gomez**; **Erway**, 2013). Nuestra investigación ha consistido en comprobar los requerimientos tecnológicos mínimos para crear un departamento de análisis forense y diseñar un modelo que sea aplicable por un número elevado de bibliotecas.

Para ello se plantean dos escenarios para bibliotecas que deseen implementar una unidad digital forense:

- el primero, de nivel básico, no necesita grandes gastos e incluso se puede poner en marcha con pocos accesorios;
- el segundo, de nivel avanzado, requiere la compra de licencias de software comercial y de hardware especializado en análisis forense, que analizaremos en detalle.

En el nivel básico es posible utilizar software gratuito perfectamente válido para los objetivos de preservación digital, como:

- *suite* de análisis forense *Autopsy*;
- creador de imágenes de disco *FTK Imager*;
- identificador de ficheros *Droid*;
- generador de valores *checksum MD5summer*;
- validador de formatos de fichero *Jhove*.

Existen dos opciones para hardware:

- Adquirir una estación de trabajo con las características adecuadas. Debería tener instaladas dos unidades de disquete (una de 5,25 pulgadas y otra de 3,5 pulgadas), capacidad para conectar un mínimo de cuatro unidades bajo la interfaz SATA, unidad externa de disquete Zip, lector de cd-rom y dvd-rom y un *write-blocker* externo;
- Aprovechar las estaciones ya instaladas en la biblioteca y sólo adquirir el material necesario para la creación de imágenes de disco. En este caso se deben adquirir disqueteras externas e internas para disquetes de 5,25, 3,5 y Zip junto con tarjetas controladoras necesarias para su conexión y un equipo *write-blocker.*

Figura 2. Torre de análisis forense creada a medida a partir de varios componentes

En el nivel avanzado el software se concentraría principalmente en la compra de una *suite* comercial de análisis forense que integre todas las funciones necesarias para la unidad. Las dos más importantes del mercado son *EnCase Forensic*, de la empresa *Guidance Software*, y *Forensic Toolkit* (*FTK*), de *AccessData*. La opción más adecuada de hardware es una unidad *FRED*, un sistema especialmente diseñado y optimizado para el análisis forense.

A continuación pasamos a describir someramente estos componentes.

4.1. Software

***FTK Imager* (*AccessData*)**
http://www.accessdata.com

El primer programa necesario es el de creación de imágenes de disco. *FTK Imager* es un software gratuito que permite trabajar con prácticamente cualquier tipo de soporte y tiene la función añadida de generar valores *checksum* de los tipos MD5 y SHA-1. Es posible acceder a los metadatos de los ficheros con la opción *Properties* y también visualizarlos mediante *File list* si se trata de ficheros pdf o bien texto simple codificado en ascii. Se puede utilizar un visor hexadecimal, que permite realizar búsquedas por texto, o explorar el contenido, lo cual es útil para localizar datos sensibles.

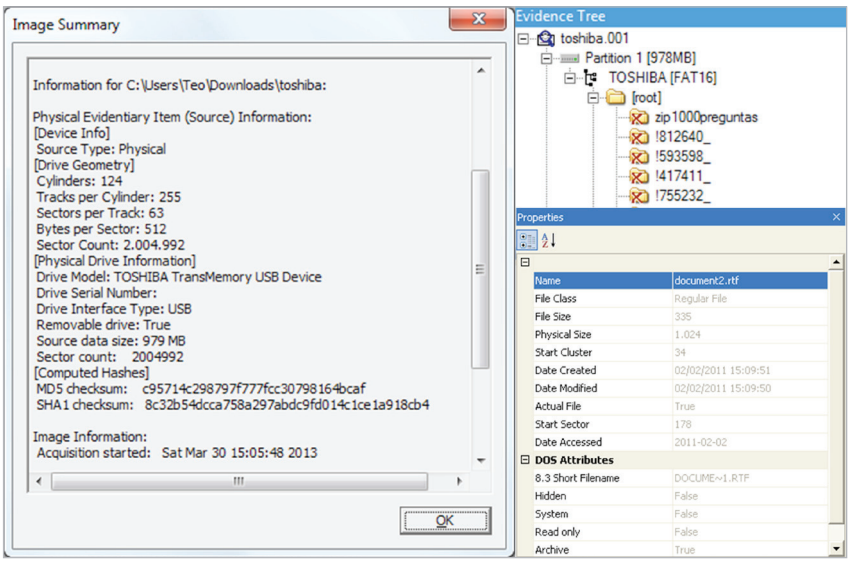

Figura 3. *FTK Imager*

MD5summer
http://www.md5summer.org

La generación de *checksums* es una función necesaria para verificar la integridad de los datos y asegurar que no se ha modificado el material original. Un programa que se adecúa perfectamente y además es de código abierto y gratuito es *MD5summer*, que además permite trabajar con múltiples ficheros y carpetas de forma recursiva. Si se desea, los resultados se pueden guardar en un fichero con la extensión md5. Este programa es realmente útil y rápido, y su única desventaja es que no puede generar valores MD5 y SHA-1 al mismo tiempo.

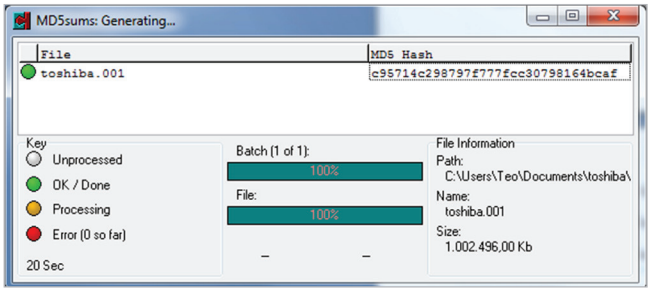

Figura 4. Generación de *checksums* en *MD5summer*

Droid (The National Archives)
http://droid.sourceforge.net

Programa desarrollado en *Java* en código libre por *The National Archives* (*TNA*) del Reino Unido que permite identificar más de novecientos formatos de fichero según el esquema PUID (*pronom's persistent unique identifier*). Su uso facilita la rápida identificación de los formatos de los ficheros que se habrán de almacenar y ofrecer a los usuarios.

Jhove (Jstor y Harvard University Library)
http://sourceforge.net/projects/jhove

La validación de los ficheros es otra tarea vital, dado que la extensión de un fichero no significa necesariamente que el contenido sea consistente con su formato. Para ello, tenemos el programa libre *Jhove* desarrollado en *Java* por la *Harvard University Library* junto con la biblioteca digital *Jstor*. Aunque el programa reconoce un número limitado de formatos, es útil para validar objetos digitales y extraer metadatos de ficheros de audio aiff y wav, de imagen tiff, jpg y gif, documentos en pdf, xml y html, y texto codificado en ascii y utf. En todo caso, es posible guardar la información de los metadatos en un fichero de texto o en documento xml.

FTK (AccessData)
http://accessdata.com/solutions/digital-forensics/forensic-toolkit-ftk

Encase forensic (Guidance Software)
http://www.guidancesoftware.com/products/Pages/encase-forensic/overview.aspx

Autopsy (The Sleuth Kit)
http://www.sleuthkit.org

Las *suites* de análisis digital forense hacen todo el proceso con un solo programa. Las más potentes y fiables son las comerciales *FTK* y *Encase Forensic*. Su uso es aceptado para la creación de evidencias digitales admitidas en tribunales de justicia (**Mercuri**, 2010). Otra opción es la gratuita en código abierto *Autopsy*, que forma parte de la biblioteca de herramientas forenses *The Sleuth Kit*. Permite analizar tanto imágenes de disco como ficheros y carpetas sueltos, además de poder recuperar ficheros "huérfanos" o borrados accidentalmente. Además, extrae metadatos de ficheros jpg y genera informes de resultados en formato html y *Excel*.

4.2. Hardware

***FRED* (*Digital Intelligence*)**
http://www.digitalintelligence.com/products/fred

Las unidades *FRED* están optimizadas para la adquisición y el análisis de datos y facilitan la lectura y duplicación de prácticamente cualquier dispositivo sin peligro de alterar los datos originales. Permiten la conexión directa de discos duros internos, arranque dual de sistemas operativos con la opción de instalar *Linux* y conexión a red por interfaz *Ethernet* que posibilita su uso como estación de trabajo estándar. La adquisición de una unidad incluye software de creación de imágenes de disco, antivirus y de análisis forense.

Figura 5. Unidad *FRED*.
(Foto de *Digital Intelligence*)

***KryoFlux* (*Software Preservation Society*)**
http://kryoflux.com

Es una tarjeta controladora de unidades de disquete de 5,25 y 3,5 pulgadas diseñada especialmente para la preservación digital. Recordemos que las placas base de los actuales ordenadores no reconocen las antiguas disqueteras de 5,25 pulgadas, por lo que estas sólo se pueden conectar mediante una tarjeta *KryoFlux* o soluciones más complejas. *KryoFlux* lee cualquier tipo de formato de disquete, aunque esté codificado o defectuoso, mediante el sistema de lectura de flujo magnético. Incluye *DTC* (*DiskTool Console*), un software específico para la creación de imágenes, que luego podrán ser procesadas con herramientas como *FTK Imager* o *Autopsy*. Una ventaja de los actuales modelos es que pueden bloquear la escritura de disquetes quitando uno de los *jumpers* de la placa.

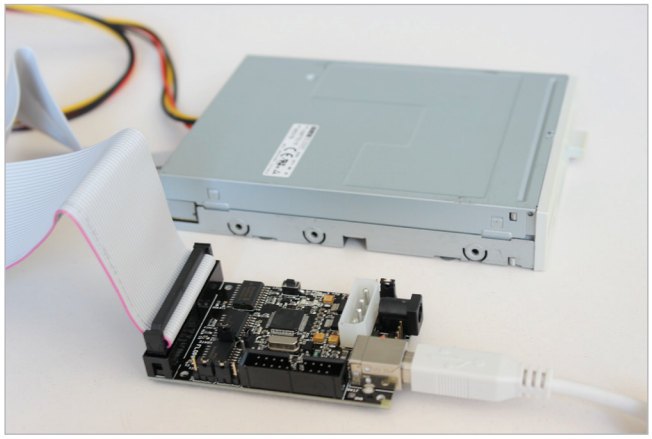

Figura 6. Tarjeta *KryoFlux* conectada a una disquetera de 3,5"

Unidades de disquete

La compra de unidades lectoras de disquete también es necesaria. En el caso de la unidad de 5,25 pulgadas, se debería adquirir una que permita lectura y escritura de disquetes con los tipos de formateado más habituales, el de 360 KB y el de 1,2 MB. El modelo más recomendado por su fiabilidad es *TEAC FD-55GFR*, del que aún se pueden encontrar unidades de segunda mano en portales como *eBay*. Por otro lado, la adquisición de unidades de 3,5 pulgadas es mucho más sencilla, ya que es fácil encontrar unidades externas con conexión por USB, aunque hay que tener en cuenta el triple formateado en estos disquetes en los PCs (360 KB, 720 KB y 1,44 MB), que no siempre es soportado en todas las unidades lectoras. Finalmente, también se debería considerar la adquisición de unidad de disquete Zip, que también presenta múltiples capacidades (100, 250 y 750 MB). En este caso, se encuentra disponible en el mercado una unidad externa de la casa *Iomega* con conexión USB 2.0 y compatible con todas las versiones de este estándar.

Docking stations

Es posible utilizar discos duros internos como si fueran externos gracias a estos dispositivos que se conectan al ordenador mediante conexión USB. Existen distintos modelos en el mercado, pero lo ideal es que dispongan de doble conexión IDE y SATA ya que son las interfaces comunes en los últimos años para discos duros, y que además sean compatibles para los de 3,5 pulgadas (usados en ordenadores de sobremesa) y 2,5 pulgadas (usados en ordenadores portátiles).

Figura 8. *Docking* station con un disco SATA insertado

Write-blockers

Para poder realizar el análisis forense es imprescindible asegurar que los datos de un disco duro o dispositivo USB no se alteren o modifiquen cuando se conecta a la unidad. Para ello existen los equipos *write-blockers*, que consiguen que los dispositivos conectados a los mismos –en especial discos duros– pasen a ser de sólo lectura, como ya lo son de origen los CD-ROM y DVD-ROM. Existen muchos modelos en el mercado, pero el más adecuado sería uno compatible con interfaces IDE, USB, SATA y opcionalmente con SCSI y *FireWire* (estándares antiguos de transferencia de datos).

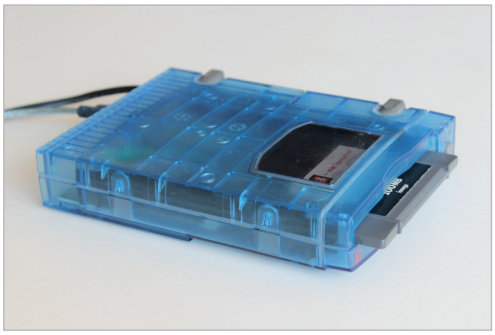

Figura 7. Unidad Zip con conexión USB

Figura 9. Duplicador de discos con sistema write-bloqued. En este caso está realizando una copia fidedigna del contenido de un antiguo disco duro IDE en un nuevo disco duro SATA

4.3. Procedimientos de trabajo

El trabajo con materiales digitales obsoletos empieza antes de su recepción física, pues es muy importante contar con el respaldo legal desde el momento de su donación, así como obtener la máxima información posible sobre sus características técnicas, y el entorno en el que se originaron (**Redwine *et al.***, 2013). Luego empieza la fase de tratamiento forense que habrá de posibilitar el acceso a los contenidos. Basándonos en los métodos y herramientas que se utilizan en las unidades presentes en los centros del apartado 3 y en experiencias relevantes de formación (**Lee**; **Woods**, 2011), se pueden diferenciar dos tipos de flujo de trabajo: uno para disquetes y otro para discos duros, ya que las capacidades y las formas de extracción de los datos presentan grandes diferencias.

> Las técnicas de análisis forense que han popularizado series televisivas como CSI se pueden aplicar en las bibliotecas para la recuperación de documentos digitales

Para disquetes, una vez adquiridos los materiales, se debe documentar su descripción física y su sistema de ficheros mediante metadatos junto con una fotografía de cada soporte. El sistema de ficheros utilizado es una información vital, ya que la existencia de diversos estándares informáticos ha propiciado la aparición de múltiples sistemas a lo largo del tiempo y su no correcta identificación podría bloquear el acceso a los datos. Los más utilizados en España son *FAT* (para *MS-Dos* y *Windows*), *NTFS* (*Windows*), *ISO 9660* (CD-ROM y DVD-ROM), *HFS* (*Mac OS*) y *extended file system* o *ext* (*Linux*). Lo ideal es que el donante proporcione esta información, aunque también se puede deducir por la época en que se crearon los contenidos y/o por los programas que se usaron. Por ejemplo, si el donante usó un sistema *Apple*, el sistema por lógica será *HFS*. Una vez completado este paso se debe tener en cuenta que hay que evitar cualquier modificación accidental de los datos originales, pues pondríamos en cuestión su autenticidad; para ello se debe bloquear la posibilidad de escritura de los disquetes. En el caso de los soportes de 5,25 y de 3,5 pulgadas el bloqueo se puede hacer manualmente, actuando sobre la correspondiente muesca de los disquetes, aunque es recomendable utilizar también la función de bloqueo que se encuentra en los modelos más recientes de la tarjeta *Kryoflux*. Los discos Zip sólo utilizaban una protección por software, por lo que se deberá utilizar un *write-blocker* para esta función.

> La aplicación de las técnicas forenses requiere disponer de equipamiento específico y de personal con la adecuada formación

A continuación se crea la imagen de disco del dispositivo, se preservan los ficheros originales y se generan valores *checksums* de la imagen y de los ficheros. El software incluido en *Kryoflux* sólo crea la imagen de disco, así que es necesario utilizar programas complementarios como *MD5summer* para verificar el *checksum* y *FTK Imager* para extraer los

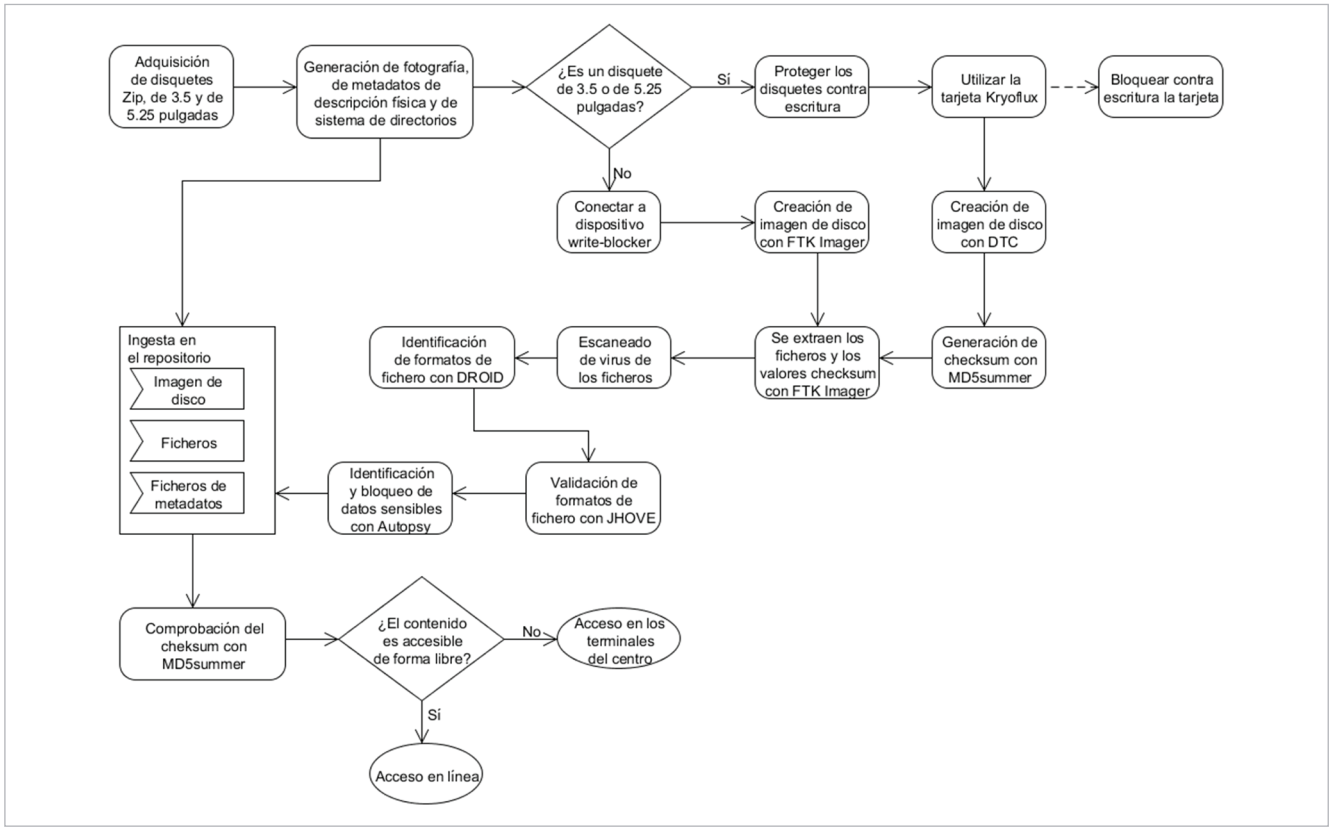

Figura 10. Workflow de preservación de disquetes

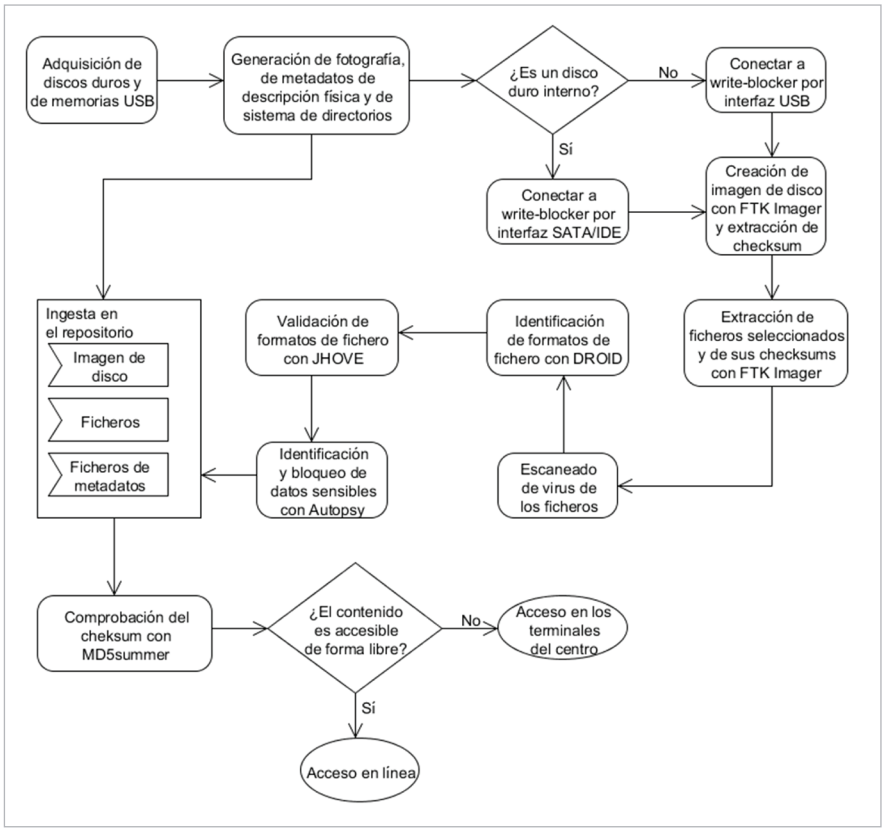

Figura 11. Workflow de preservación de discos duros y memorias USB

ficheros y los valores correspondientes en ficheros .csv. El tiempo de creación de una imagen dependerá de la cantidad de errores que presente el soporte. El software recuperará el máximo de datos posible para minimizar las pérdidas de información. Seguidamente se debe pasar un antivirus por los ficheros y la imagen para evitar la entrada de cualquier contenido *malware* en las estaciones de trabajo del centro.

La siguiente fase, una vez ya se encuentran los contenidos en las estaciones de trabajo, es identificar mediante *Droid* los formatos de fichero para así saber exactamente qué tipo de contenidos contienen los materiales, para seguidamente validarlos mediante *Jhove*. Por último es muy importante localizar la información sensible que pueda existir, como información personal, bancaria, filiación política, etc., o cualquier otra que el donante haya especificado si se da el caso, con el fin de bloquear su consulta pública y así cumplir con la legislación de protección de datos, que en España es la *Ley orgánica de protección de datos de carácter personal* (*España*, 1999). El programa adecuado sería *Autopsy*.

El procedimiento final para asegurar la preservación de los contenidos es el depósito de los ficheros de metadatos, de las imágenes de disco y de los ficheros en el repositorio interno del centro y comprobar posteriormente que se han subido los datos de forma correcta con un validador de *checksums* como *MD5summer*. Dependerá de la política del centro si los contenidos se podrán consultar libremente o su acceso será restringido a investigadores.

El proceso anterior es parecido con discos duros, pero teniendo en cuenta que la cantidad de ficheros con conexión IDE, SATA o USB es mucho más elevada que en disquetes, por lo que no es razonable hacer la extracción de todos. Es más lógico realizar una selección, ya sea porque el donante así lo haya especificado o bien porque el centro encuentra algunos contenidos de interés especial, y luego preservar la imagen de disco que contendrá los contenidos originales y los ficheros seleccionados.

El inicio del proceso de trabajo es idéntico al descrito para los disquetes, con la generación de la fotografía y los metadatos. Los soportes que se utilizan en este caso tienen activados los atributos de escritura y no se pueden bloquear directamente, por lo que es obligado el uso de sistemas *write-blocker*.

Seguidamente se realiza la creación de la imagen de disco con *FTK Imager*, que generará los informes de *checksum* correspondientes. A continuación se extraen los ficheros seleccionados y de cada uno se genera también su *checksum*. Igual que con los disquetes, hay que analizar los contenidos con un software antivirus, identificar y validar los formatos de fichero, y buscar y bloquear los datos sensibles.

El paso final es la ingesta en el repositorio de los ficheros de metadatos, los ficheros seleccionados y la imagen de disco. Es importante tener en cuenta el coste que implica el gran volumen de espacio necesario para almacenar colecciones de discos duros (**Woods**; **Lee**; **Garfinkel**, 2011).

> Es necesaria la incorporación de personal informático pero también será imprescindible que el personal bibliotecario se forme en estos nuevos temas

5. Conclusiones

La evolución de la tecnología informática presenta diversos retos a las instituciones encargadas de la preservación y la difusión del conocimiento. Uno de los principales es ser capaces de continuar dando acceso a los usuarios a datos y documentos que se originaron y conservaron en medios actualmente obsoletos. Hemos visto que las técnicas de análisis forense digital están disponibles para su uso por las bibliotecas y también que algunas de ellas ya han emprendido este camino.

No es fácil el correcto manejo de equipos de análisis forense digital, pues para ello se requiere tener conocimientos y habilidades que hasta ahora no han sido habituales entre el personal de las bibliotecas. Algunas iniciativas ya están trabajando a nivel internacional para conseguir que el currículo de los futuros profesionales de las bibliotecas también acoja este tipo de formación (**Tibbo**; **Lee**, 2012), como fue re-

cogido en la conferencia celebrada por el proyecto *DigiCur* (**Cirinnà**; **Fernie**; **Lunghi**, 2013). También se han de destacar proyectos como *BitCurator* que tienen como objetivo crear paquetes de software seleccionados y configurados para su uso específico en bibliotecas.
http://www.bitcurator.net

Aunque el montaje a nivel profesional de un laboratorio forense digital es complicado (**Jones**; **Valli**, 2008), los laboratorios que están creando las bibliotecas tienen una menor complejidad, entre otras cosas porque se enfrentan a problemas técnicos de un nivel inferior a los propios de los laboratorios de criminalística. Nuestra investigación muestra que el montaje de una unidad forense de bajo o medio coste es posible en instituciones medianas o grandes y se pueden conseguir grandes ventajas disponiendo de ella.

Es cierto que los requerimientos técnicos no son fáciles de implementar en una primera etapa pues es necesario disponer de un cierto conocimiento técnico sobre la materia. Por ello no se puede descartar que las unidades de análisis forense sean soportadas por más de una institución. Sería deseable que tuvieran un importante papel de referencia en ello bibliotecas nacionales, regionales y consorcios de bibliotecas.

Nota

1. El *checksum* es un número de control resultado de aplicar un algoritmo de cálculo a una secuencia de datos, que se transmite junto a éstos. Al recibirlos, el receptor calcula el *checksum* para verificar que no haya discrepancia con el valor inicial. Si no coincide se rechazan los datos o se pide una retransmisión. Ejemplos de *checksum* son el último número del ISBN, o la letra del NIF.

Agradecimientos

Investigación realizada dentro del proyecto *El acceso abierto (open access) a la ciencia en España*. 2012-2014. *Plan Nacional I+D+i*, código CSO2011-29503-C02-01.

6. Bibliografía

AIMS Work Group (2012). *AIMS born-digital collections: an inter-Institutional model for stewardship*.
http://www2.lib.virginia.edu/aims/whitepaper/AIMS_final_A4.pdf

Barrera-Gómez, Julianna; **Erway, Ricky** (2013). *Walk this way: Detailed steps for transferring born digital content from media you can read in house*. Ohio: OCLC. ISBN: 978 1 55653 454 6
http://www.oclc.org/content/dam/research/publications/library/2013/2013-02.pdf

Carroll, Laura; **Farr, Erika**; **Hornsby, Peter**; **Ranker, Ben** (2011). "A comprehensive approach to born-digital archives". *Archivaria*, v. 72, pp. 61-92.
http://pid.emory.edu/ark:/25593/cksgv

Cirinnà, Chiara; **Fernie, Kate**; **Lunghi, Maurizio** (2013). *Proceedings of the Framing the Digital Curation Curriculum conference* (Florencia, Italia, 6-7 mayo 2013).
http://www.digcur-education.org/eng/International-Conference/DigCurV-2013-proceedings

Del-Pozo, Nicholas; **Elford, Douglas**; **Pearson, David** (2009). "Invited demo: Prometheus: managing the ingest of media carriers". *DigCCurr 2009*, pp. 73-75.
http://www.nla.gov.au/openpublish/index.php/nlasp/article/view/1384/1674

Edwards, Glynn; **Chan, Peter**; **Olson, Michael** (2010). *First draft of our forensic workflow*. Stanford University.
http://lib.stanford.edu/digital-forensics-stanford-university-libraries/first-draft-our-forensic-workflow

Erway, Ricky (2012). *You've got to walk before you can run first steps for managing born digital content received on physical media*. Ohio: OCLC.
http://www.oclc.org/content/dam/research/publications/library/2012/2012-06.pdf

España (1999). "Ley orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal". *BOE*, n. 298, 14 de diciembre.
https://www.boe.es/buscar/act.php?id=BOE-A-1999-23750

John, Jeremy-Leighton (2012). *Digital forensics and preservation.* Digital Preservation Coalition. Technology watch report, 12-03. Hesligton: Digital Preservation Coalition.
http://www.dpconline.org/component/docman/doc_download/810-dpctw12-03pdf
http://dx.doi.org/10.7207/twr12-03

Jones, Andrew; **Valli, Craig** (2008). *Building a digital forensic laboratory: Establishing and managing a successful facility*. Burlington (MA, EUA): Elsevier. ISBN: 9781856175104

Kirschenbaum, Matthew G.; **Ovenden, Richard**; **Redwine, Gabriela** (2010). *Digital forensics and born-digital content in cultural heritage collections*. Washington, DC: Council on Library and Information Resources. ISBN: 978 1 932326 37 6
http://www.clir.org/pubs/reports/pub149/reports/pub149/pub149.pdf

Laing, Gordon (2004). *Digital retro: the evolution and design of the personal computer.* Alameda, CA: Sybex, ISBN: 078214330X

Lee, Christopher A.; **Woods, Kam** (2011). *Digital acquisition learning laboratory: A white paper*. School of Information and Library Science. University of North Carolina at Chapel Hill.
http://www.ils.unc.edu/callee/dall-white-paper.pdf

Loftus, Mary J. (2010). "The author's desktop". *Emory magazine*, Winter, pp. 22-27.
https://www.emory.edu/EMORY_MAGAZINE/2010/winter/winter-2010.pdf

Mercuri, Rebecca (2010). "Criminal defense challenges in computer forensics". *Digital forensics and cyber-crime*. Berlin; Heidelberg: Springer-Verlag, pp. 132-138. ISBN: 9783642115332
http://goo.gl/aQlpdv
http://dx.doi.org/10.1007/978-3-642-11534-9_13

Olson, Michael (2011). "The Stanford Forensics Lab: A case study". *Digital forensics for preservation.*
http://www.dpconline.org/component/docman/doc_download/628-forensicsolson

Palmer, Gary (2001). *A road map for digital forensic research*. Utica, NY: Air Force Research Laboratory, Rome Research Site. *http://www.dfrws.org/2001/dfrws-rm-final.pdf*

Paradigm Project (2007). *Workbook on digital private papers*. *http://www.paradigm.ac.uk/workbook*

Redwine, Gabriella; **Barnard, Megan**; **Donovan, Kate**; **Farr, Erika**; **Forstrom, Michael**; **Hansen, Will**; **Leighton, John**; **Kuhl, Nancy**; **Shaw, Seth**; **Thomas, Susan** (2013). *Born digital: Guidance for donors, dealers, and archival repositories*. Washington DC: Council on Library and Information Resources. ISBN: 9781932326468
http://www.clir.org/pubs/reports/pub159/pub159.pdf

Thomas, Susan (2011). "Curating the I, digital: experiences at the Bodleian Library". En: Lee, Christopher A. *I, digital: personal collections in the digital era.* Chicago: Society of American Archivists, pp. 280-301. ISBN: 1931666385

Tibbo, Helen R.; **Lee, Christopher A.** (2012). "Closing the digital curation gap: A grounded framework for providing guidance and education in digital curation". En: *Proceedings of Archiving 2012*, pp. 57-62.
http://ils.unc.edu/callee/p57-tibbo.pdf

Verheul, Ingeborg (2006). *Networking for digital preservation: current practice in 15 national libraries*. München: Saur. ISBN: 3598218478
http://www.ifla.org/files/assets/hq/publications/ifla-publications-series-119.pdf

Woods, Kam; **Lee, Christopher A.**; **Garfinkel, Simson** (2011). "Extending digital repository architectures to support disk image preservation and access". En: *Procs of the 11th Annual Intl ACM/IEEE joint conf on digital libraries JCDL'11*, pp. 57-66.
http://www.ils.unc.edu/callee/p57-woods.pdf
http://dx.doi.org/10.1145/1998076.1998088