

UC Irvine

UC Irvine Previously Published Works

Title

Hacia Flujos de Trabajo Responsables y Sostenibles con la Creación de Imágenes de Disco Digitales

Permalink

<https://escholarship.org/uc/item/2d89s3wz>

Author

Arroyo-Ramirez, Elvia

Publication Date

2021-07-27

HACIA FLUJOS DE TRABAJO RESPONSABLES Y SOSTENIBLES CON LA CREACIÓN DE IMÁGENES DE DISCO DIGITALES¹

Elvia Arroyo-Ramírez
Asistente Archivera Universitaria
Departamento de Colecciones Especiales y Archivos
Biblioteca de la Universidad de California, Irvine (UC Irvine)

Prepared for delivery at the **2021 Virtual Congress** of the Latin American Studies Association, **May 26–29, 2021**

Área temática: SCH / Scholarly Resources
Panel: “Historia digital: trabajando con archivos y tecnologías digitales”

Resumen:

En años recientes (circa 2016-presente) erudición y práctica de la comunidad archivera ha visto incrementada la crítica sobre la práctica de la creación de imagen de disco por defecto debido a cuestiones sobre los problemas éticos, laborales, y la sostenibilidad que surgen al crear estas imágenes. La comunidad de archiveros en los Estados Unidos y Canadá ha empezado a modificar las mejores prácticas para proponer flujos de trabajo alternativos que no predeterminan o dependen de la creación de imágenes de disco y animan intervenciones curatoriales tempranas como la evaluación de materiales antes que lleguen al repositorio.

Esta ponencia se informa en el caso del Departamento de Colecciones Especiales y Archivos en la Universidad de California, Irvine (UCI) y cómo este cambio ha afectado las prioridades y los flujos de trabajo, en particular durante este momento de trabajo distanciado debido a COVID-19. Voy a presentar un nuevo flujo de trabajo que permite al personal seguir procesando y preservando colecciones desde su casa con herramientas ligeras y de fuente abierta (“open-source”) y sin necesitar herramientas de costo o hardware.

Keywords: archives, born-digital, born-digital workflows, digital preservation, disk-imaging, ethics in digital archives, remote work, open-source software, COVID-19 pandemic

¹ Muchas gracias a mis colegas Nicolas Quiroga, David Bliss, y Denise Frigo con quien presente esta ponencia durante nuestro panel titulado, “Historia digital: trabajando con archivos nacidos digitales.” Gracias especiales a Eden Molina quien fue nuestra comentadora durante el panel. Gracias especiales a Nicolas Quiroga y Lourdes Johnson quienes dieron comentarios editoriales.

Introducción

El trabajo de archiveros está conectado inextricablemente a los materiales efímeros como a los físicos. Al imaginar este trabajo, se puede pensar en cajas pesadas localizadas en cuartos oscuros sin luz natural, llenas de papeles en carpetas. Para preparar una colección de archivos para el acceso público se requiere trabajo físico y en persona. Al leer descripciones de solicitud de trabajo para archiveros es típico encontrar el requisito de poder levantar 18 kilos o 40 libras.

Sin embargo, en las últimas dos décadas los materiales efímeros de los archivos han incluido la prevalencia creciente de la adquisición y preservación de formatos nacidos digitales (born-digital). Archivos en estos formatos incluyen disquetes, memorias de USB, CDs, DVDs, y directorios lógicos que necesitan herramientas y flujos de trabajo especializados para extraer, analizar, preservar, y preparar archivos para preservación infinita y acceso al público. Desde la primera época (circa 2002-2012) en que la profesión notó la aparición de estos formatos comenzaron a desarrollarse mejores prácticas para preservar archivos nacidos digitales. La práctica de crear imágenes de disco (disk images) fue la práctica más aceptada y común para preservar copias del original, bit por bit. Además, muchos de los primeros flujos de trabajo para procesar materiales nacidos digitales dependían de herramientas y hardware físico que requerían trabajo en la oficina donde estaba localizado el equipo especializado para completar ese trabajo. Este fue así porque la mayoría de los primeros formatos digitales encontrados en colecciones de archivo circularon en medios físicos como disquetes, CDs, y DVDs—medios populares en uso en los años 1970s hasta la primera década de los 2000s. Para acceder al contenido de un disquete de 3,5,” por ejemplo, el personal debe tener acceso no solo al hardware, un lector de disquete que hoy se necesita comprar especialmente, sino también tener acceso a software especial para

completar la primera acción hacia la preservación que es la búsqueda de virus en el disquete. Guymager,² por ejemplo, es una herramienta disponible para hacer este trabajo.

En años recientes (circa 2016-presente) la erudición y la práctica de la comunidad profesional ha sido cada vez más crítica sobre la práctica de la creación de imagen de disco. En la siguiente sección, proporcionaré una revisión de la literatura crítica sobre los problemas éticos que se desprenden de la creación de imágenes de disco. Gracias a esos debates, la comunidad de archiveros en los Estados Unidos y Canadá ha empezado a adoptar mejores prácticas. Estas prácticas proponen flujos de trabajo alternativos que no primariamente dependen de la creación de imágenes de disco y animan intervenciones curatoriales tempranas como la evaluación de materiales antes de que lleguen al repositorio.

“Preserva ahora, analice después”: hacia prácticas de automatización a través de la imagen de disco por defecto

Las definiciones para la imagen de disco son consistentes a través de la literatura existente. En 2008, Matthew Kirschenbaum definió la imagen de disco como,

“una representación literal de cada bit de información de alguna instancia de un soporte medial original... no se trata simplemente de una copia de todos los archivos que había en ese disquete original, sino que la imagen del disco, como el facsímil o la fotografía que sugiere el término, conserva toda la información que estaba grabada en el disco en su geometría de almacenamiento original.”³

En un artículo fundamental, “Extending digital repository architectures to support disk image preservation and access,” (2011) Kam Woods, Cal Lee, y Simson Garfinkel definieron la imagen de disco como,

² <https://guymager.sourceforge.io/>.

³ Matthew Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination* (Cambridge, MA: MIT Press, 2008), 115.

“una copia sector por sector de los datos que se almacenaron en un medio físico. Como tal, la imagen del disco es una "instantánea" del contenido del medio, incluidos todos los archivos asignados, nombres de archivo y otra información de metadatos asociada con el volumen del disco.”⁴

Woods, Lee y Garfinkel consideran que las imágenes de disco "proporcionan mecanismos a prueba de fallos" (fail-safe) y "ofrecen opciones para análisis futuros" que "permiten el acceso a datos potencialmente valiosos que residen por debajo del nivel del sistema de archivos.”⁵

Monique Lassere y Jess Whyte relatan que,

“el propósito del modelo de imágenes de disco es que el investigador, el archivero, el curador o el bibliotecario siempre pueden volver a la imagen del disco original para aplazar el trabajo adicional, demostrar autenticidad, permitir la emulación o el acceso, o generar nuevas copias de acceso.”⁶

De hecho, para la primera generación de archiveros que aprendieron a trabajar con archivos digitales de esa época, crear imágenes de disco presentó una opción fácil y eficiente para capturar no solo el contenido de los archivos sino también datos sobre el soporte y el sistema que se usó cuando los archivos fueron creados. Fue así como el primer modelo de preservación vía captura de imagen de disco se instaló como la mejor práctica en flujos de trabajo y la actitud de “preserva ahora, analice después” fue común en la preservación de archivos nacidos digitales.

El desarrollo del modelo de “preserva ahora, analice después” está de acuerdo con la tendencia hacia las prácticas de automatización en la curación y el análisis digital forense. Es muy probable que la tendencia hacia la automatización se deba a la casi segura realidad de que la mayoría de los repositorios de archivos están plagados de materiales sin procesar o inaccesibles;

⁴ Kam Woods, Cal Lee, y Simson Garfinkel, “Extending Digital Repository Architectures to Support Disk Image Preservation and Access,” en *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries* (New York: Association of Computing Machinery, Junio 2011), 64.

⁵ Woods, Lee, Garfinkel, 64

⁶ Monique Lassere y Jess Whyte, “Balancing Care and Authenticity in Digital Collections: A Radical Empathy Approach to Working with Disk Images,” in “Radical Empathy in Archival Practice,” eds. Elvia Arroyo-Ramirez, Jasmine Jones, Shannon O’Neill, and Holly Smith. Special issue, *Journal of Critical Library and Information Studies* 3, 4.

la automatización puede ayudar a acelerar la capacidad de los archiveros para reducir sus atrasos. Pero como ya he comentado en otro artículo, la automatización no siempre es una estrategia ganadora, ya que puede provocar pérdidas involuntarias como pérdida de contexto y significado por eliminación errónea de marcas diacríticas.⁷

Otra realidad importante que se debe abordar es la actitud similar como es la de “tómalo y guárdalo todo”, en gran parte resultado de la falta de roles definidos en la colaboración entre curadores y archiveros digitales y la necesidad de tomar decisiones de revisión, intervención y evaluación previas a la custodia antes de que el repositorio acepte materiales. Muchos repositorios aún no tienen los recursos adecuados como tampoco personal capacitado y roles dedicados a la archivista digital. La preservación de materiales creados en formato digital requiere una gran inversión de tiempo y recursos de personal. Muchos repositorios de archivos todavía carecen de personal específicamente capacitado para desarrollar flujos de trabajo adecuados y este sigue siendo un problema importante.

Críticas sobre práctica de la creación de imagen de disco por defecto

Uno de los problemas más significativos en crear imágenes de disco es la captura de datos sobre los que los donantes o creadores de contenido no dieron consentimiento para su disponibilidad, acaso debido a que no sabían que sus archivos borrados u ocultos aún podían ser accesibles. Esto lo experimenté de primera mano cuando procesé mi primera colección nacida digital, los papeles del poeta argentino Juan Gelman. Mientras procesaba 164 disquetes que documentaban su esfuerzo para enjuiciar a los responsables del asesinato de su hijo y nuera y la

⁷ Elvia Arroyo-Ramírez, “Invisible Defaults and Perceived Limitations: Processing the Juan Gelman Files,” (presentación, Preservation and Archiving Special Interest Group (PASIG), New York, October 28, 2016), <https://medium.com/on-archivy/invisible-defaults-and-perceived-limitations-processing-the-juan-gelman-files-4187fdd36759>.

desaparición de su nieta, pude crear imágenes de disco de cada disquete. Con estas imágenes de disco, logré hacer un análisis más detallado utilizando la herramienta *Disk Image Access* que está disponible como parte del conjunto de herramientas *BitCurator*. Con esta herramienta, los archiveros no solo pueden ver una lista de directorios de todos los nombres de archivos que están escritos actualmente en el disco, sino también los nombres de los archivos que han sido eliminados o están ocultos y que aparecen en texto rojo. Especulo que Gelman, como muchos usuarios de esta época de la creación de archivos digitales (circa 1990s-2000s), creía que estos archivos, una vez eliminados u ocultos, serían inaccesibles para los futuros usuarios de su archivo. Aunque no capture la pantalla de un ejemplo del caso de Juan Gelman, *imagen 1* abajo es un ejemplo de cómo se muestran estos archivos eliminados u ocultos en la herramienta *Disk Image Access*. Adicionalmente, herramientas como *Disk Image Access* son herramientas que se utilizan exclusivamente con imágenes de disco, haciendo el uso de imágenes de disco esenciales para seguir utilizando ciertas herramientas.

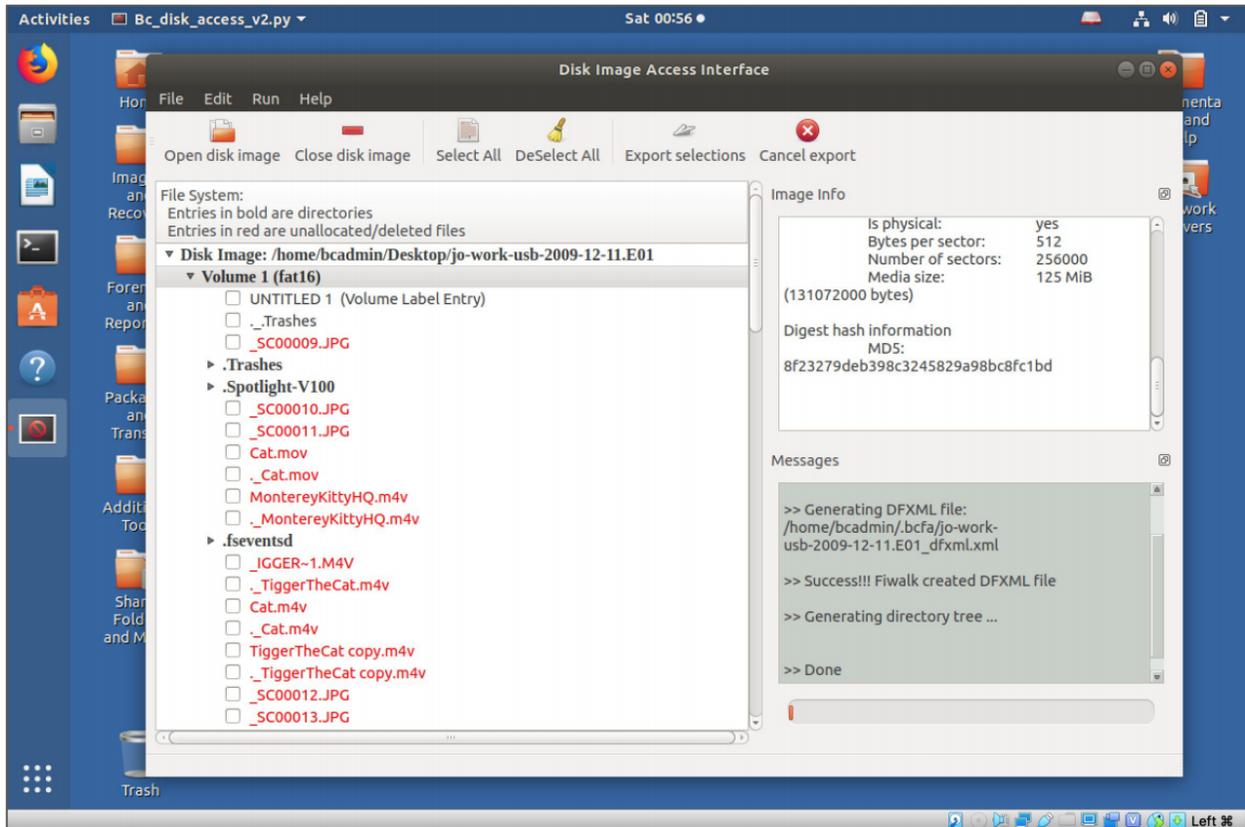


Imagen 1. Captura de pantalla en el ambiente de BitCurator utilizando la herramienta Disk Image Access. Los nombres de archivo que aparecen en rojo son archivos borrados por el donante o creador. Imagen tomada del BitCurator Quick Start Guide, versión actual: <http://distro.ibiblio.org/bitcurator/docs/BitCurator-Quickstart-v2.2.pdf>.

En los comienzos de la admisión de formatos nacidos digitales por parte de los repositorios, no estaba claro cómo preservar y acceder a esos archivos. En la archivística, muchos han tratado este tema, y en el 2020, Monique Lassere y Jess Whyte en su artículo “Balancing Care and Authenticity in Digital Collections: A Radical Empathy Approach to Working with Disk Images” sostuvieron que crear imágenes de disco es una brecha de confianza y “falta de atención a la privacidad y seguridad de los creadores o sujetos” que aparecen en los archivos de otros.⁸ En el caso de sujetos, o donantes o creadores de archivos ya difuntos, como

⁸ Lassere y Whyte, 11.

en el caso del archivo Gelman, donado por la viuda del escritor a la biblioteca de Firestone de la Universidad de Princeton, estos individuos ni siquiera pueden ser consultados para dar su consentimiento. Lassere y Whyte relatan que estas decisiones son laboriosas y recomiendan tomar decisiones empáticas, contrarias a las decisiones basadas en automatización, cuando se trata de determinar si la creación de imágenes de disco representaría riesgos o daños para los creadores, donantes y sujetos de los archivos.

Lassere y Whyte también abordan el impacto que las prácticas de imagen de disco por defecto tienen sobre el trabajo y los recursos, y abogan por la evaluación y las intervenciones previas a la custodia. La actitud de "tomar y salvar todo" es un "método de aplazamiento"⁹ que en realidad es un aplazamiento de la responsabilidad de los selectores que coloca injustamente a los archiveros digitales en un problema en el que su trabajo está implicado dos veces: primero al crear la imagen de disco y luego en los análisis posteriores. Desafortunadamente, en muchos casos el ritmo con el que los archiveros están recibiendo nuevos archivos digitales no permite que se realicen estos análisis por falta de recursos, por lo que la promesa del examen crítico del contenido de lo que se capturó es en esencia una promesa no sostenible e irresponsable. Lassere y Whyte relatan que,

“A menos que el proceso de adquisición pueda modificarse para evaluar mejor los materiales nacidos digitales, los archiveros y curadores deben limitar explícitamente las adquisiciones de materiales nacidos digitales para que estén de acuerdo con *la capacidad* del archivo para administrar éticamente estos materiales.”¹⁰ (las cursivas son mías)

Más adelante en esta ponencia, discutiré mi trabajo reciente en actualizar la guía titulada *University of California Guidelines for Efficient Archival Processing* que brinda soluciones

⁹ Lassere y Whyte, 22.

¹⁰ Lassere y Whyte, 19.

prácticas para reasignar esta responsabilidad a los curadores y selectores y fomenta la colaboración entre archiveros y curadores durante las fases de pre-custodia y evaluación.

Alternativas más éticas e sostenibles

En la primera época de preservación de colecciones nacidas digitales, las preocupaciones sobre el almacenamiento eran mínimas. Los primeros formatos que comenzaron a aparecer en los repositorios de archivo fueron los disquetes de 3,5" y 5" que sus capacidades de disco eran comparativamente pequeñas. Hoy en día, los repositorios tienen necesidades de almacenamiento cada vez mayores con adquisiciones digitales también cada vez más grandes, de cientos de gigabytes o terabytes. Crear imágenes de disco con capacidades contemporáneas en USBs o hard drives, es crear una copia exacta, que puede incluir archivos duplicados y espacio sin uso y no asignado. La imagen del disco siempre será del mismo tamaño del dispositivo de almacenamiento, no del contenido de los archivos existentes. Esta práctica por defecto de crear imágenes de disco quiere decir que ahora los repositorios han preservado grandes cantidades de espacio en disco no utilizado, archivos eliminados y otros datos que ni el donante ni el repositorio realmente quieren conservar. Esto produce impactos en varios recursos traducidos en problemas económicos, laborales y ambientales.

Temas relacionados a la sostenibilidad ambiental de la archivística digital se discuten en el artículo, “Towards Environmentally Sustainable Digital Preservation” (2019) escrito por Keith L. Pendergrass, Walker Sampson, Tessa Walsh, y Laura Alagna. Los autores analizan el impacto ambiental ya que la archivística digital depende de “infraestructura tecnológica (tecnología de la información y las comunicaciones, TIC) que implica considerables impactos ambientales

negativos.”¹¹ El artículo propone varios cambios paradigmáticos incluyendo en el área de evaluación (appraisal) para aplicar un modelo de práctica más sostenible. Ellos notan,

“En lugar de retener imágenes de disco como una copia de preservación de forma predeterminada, las organizaciones deben evaluar el impacto ambiental y financiero de esta decisión y crear pólizas y procedimientos para la eliminación y destrucción.”¹²

Los autores también discuten un supuesto sobre los usuarios que quieren tener acceso a los datos.

Es muy raro encontrar un usuario que esté interesado en los datos forenses como el sistema operativo que un creador de archivos utilizó al escribir sus archivos. Mi experiencia personal indica que el mayor número de usuarios está primariamente interesado en el contenido del archivo; y la información forense resulta demasiada o inútil para los intereses de búsqueda de los usuarios. Adicionalmente, el formato de una imagen de disco (.aff, .E01, etc.) causa confusión para usuarios no familiarizados sobre cómo acceder a esos formatos y requiere más inversión de tiempo del personal para ayudar a usuarios a encontrar las herramientas o emuladores adecuados para acceder el contenido de la imagen. Estoy de acuerdo con los autores cuando escriben que sí “existen casos de retención a largo plazo de imágenes de disco forenses completas, pero [estos casos] pueden ser muchos menos de lo que comúnmente se concibe.”¹³ En la sección final de esta ponencia compartiré un caso de uso en el que sí es apropiado crear imágenes de disco para preservación y acceso.

Grupo DANNG

El grupo estadounidense DANNG (Digital Archival traNsfer, iNgest, and packagiNg Group) actualmente está desarrollando una guía titulada, “Disk Imaging Decision Factors” donde

¹¹ Keith L. Pendergrass, Walker Sampson, Tessa Walsh, y Laura Alagna, “Towards Environmentally Sustainable Digital Preservation” en *The American Archivist* 82 (1): 165.

¹² Pendergrass, Sampson, Walsh, y Alagna, 183.

¹³ Pendergrass, Sampson, Walsh, y Alagna, 183.

críticamente exploran criterios que ayuden a los archiveros a decidir cuándo es apropiado crear imágenes de disco y cuando no es necesario. Unos de los criterios mencionados, “¿es la interactividad una función deseada para los usuarios?”¹⁴ ayuda a fomentar casos en donde la imagen de disco es quizás esencial. La interactividad es una importante función en contenidos más complejos que archivos de texto o fotografía como, por ejemplo, software propietario, menús de DVD, y ambientes a los que solo se puede acceder con la emulación. Los criterios para no crear imágenes de disco incluyen consideraciones ya mencionadas aquí como, “¿ha hablado usted o su donante sobre la pérdida y retención de datos?”¹⁵ que aborda la responsabilidad del personal de ser transparente y tener conversaciones específicas sobre los archivos digitales con donantes en los momentos de la pre-custodia. El Grupo DANNNG reconoce que la definición de las imágenes de disco debe indicar exactamente qué se preserva y accede cuando los repositorios crean imágenes de disco: “El contenido... incluye archivos visibles, pero también... archivos ocultos e incluso archivos eliminados; puede incluir cosas que el usuario puede no darse cuenta de que están guardadas en sus discos rígidos, como el correo electrónico descargado y el historial de navegación web. También puede incluir espacio no utilizado, vacío.”¹⁶ Quizás una mejor práctica sería compartir esta definición con los donantes para que ellos estén informados sobre la técnica y la práctica que está disponible y para que puedan dar consentimiento informado sobre su donación.

Otra pregunta “¿tiene un dispositivo de almacenamiento que no necesita, no puede, o probablemente no debería, copiar?” aborda el hecho de que no todos los dispositivos de almacenamiento requieren una imagen; por ejemplo, el uso del dispositivo “solo para la

¹⁴ Digital Archival Transfer, Ingest, and Packaging Group (DANNNG), “Disk Imaging Decision Factors,” version 0.3, <https://bit.ly/3uY3RB1>, 6.

¹⁵ Digital Archival Transfer, Ingest, and Packaging Group (DANNNG), “Disk Imaging Decision Factors,” version 0.3, <https://bit.ly/3uY3RB1>, 9.

¹⁶ Digital Archival Transfer, Ingest, and Packaging Group (DANNNG), “Disk Imaging Decision Factors,” version 0.3, <https://bit.ly/3uY3RB1> 3.

transferencia y no como una unidad en funcionamiento, es menos probable que tenga un "valor artificial."¹⁷ Adicionalmente, los materiales de archivo "puramente digitales" constituyen un porcentaje creciente de nuevas adquisiciones y eventualmente constituirán la mayoría de las adquisiciones digitales de archivo a medida que disminuye la cesión de contenido basado en dispositivos de almacenamiento. Actualmente la archivística tiene que reconocer que los materiales contemporáneos nacidos digitales no requieren captura de imagen de disco. Lassere y Whyte notan,

"A medida que la creación de contenido contemporáneo pasa del almacenamiento de un solo medio, como los discos duros locales, al almacenamiento distribuido y redundante (por ejemplo, un Google Drive personal), la presencia de contenido digital vinculado a los medios físicos en una colección personal será menos frecuente."¹⁸

De "mejores prácticas" hacia "prácticas suficientemente buenas": University of California Guidelines for Efficient Archival Processing

En 2020, un grupo de archiveros de la Universidad de California (UC) publicaron una nueva versión de *University of California Guidelines for Efficient Archival Processing*¹⁹ ya que tenía ocho años sin actualizar y carecía de una guía fundamental sobre cómo procesar de manera eficiente los materiales de archivo digitales. Los nuevos cambios tomaron los *Guidelines* en una dirección holística que reconoce que la prácticas de procesar archivos de manera eficiente para reducir los atrasos de materiales no procesados o inaccesibles son impactadas por problemas acerca de coleccionar, tomar colecciones sin evaluación previa (appraisal), y la accesión

¹⁷ Digital Archival Transfer, Ingest, and Packaging Group (DANNNG), "Disk Imaging Decision Factors," version 0.3, <https://bit.ly/3uY3RB1> 11.

¹⁸Lassere y Whyte, 4.

¹⁹ Kate Dundon, Laurel McPhee, Elvia Arroyo-Ramirez, et al. (2020). Guidelines for Efficient Archival Processing in the University of California Libraries (Version 4). *UC Office of the President: University of California Systemwide Libraries*. Retrieved from <https://escholarship.org/uc/item/4b81g01z>.

(accessioning).²⁰ En la versión actual de los *Guidelines* estos temas están indicados como principales causas del crecimiento de nuestros retrasos físicos y digitales y las recomendaciones —que implican a todo personal, como administradores, curadores/coleccionistas, archiveros—, se enfocan hacia mejores y más responsables prácticas. En relación con el tema de archivos digitales, los autores recomiendan varias intervenciones en las áreas de evaluación (ver sección 3) y recomiendan crear pólizas y flujos de trabajo que animen intervenciones curatoriales tempranas como la evaluación de materiales antes de que lleguen al repositorio, y antes de que se depositen permanentemente. En relación con la creación de imágenes de disco, los *Guidelines* recomiendan evitar la creación de imágenes de disco al menos que sea necesario y recomiendan la transferencia de archivos lógicos.²¹

En el área de acceso (accessioning), los *Guidelines* recomiendan un conjunto de pasos básicos (baseline accessioning) que esencialmente preparan a los archivos digitales para el acceso si no presentan riesgos de privacidad que requieran procesamiento más completo y pesado (ver sección 4.H.). Las recomendaciones utilizadas en la acceso adoptan el mismo criterio que las recomendaciones originales de los *Guidelines* que animan la adopción de “prácticas suficientemente buenas” que son distintas de las “mejores prácticas”, las que a veces son muy difíciles de implementar.

En la próxima sección relataré un flujo de trabajo completamente nuevo en la Universidad de California, Irvine (UCI) que fue influenciado directamente por la revisión de los *Guidelines* y es un flujo de trabajo de acceso básico específico para materiales que se pueden transferir a través de directorios lógicos.

²⁰ Elvia Arroyo-Ramirez, Kate Dundon, y Shira Peltzman, “Holistic approaches to born-digital appraisal and accessioning—revising the UC Guidelines for Efficient Archival Processing,” presentación en *OCLC Works in Progress* webinar, fecha 16 Marzo 2021, <https://escholarship.org/uc/item/7k45c35j>.

²¹ Guidelines for Efficient Archival Processing in the University of California Libraries (Version 4), 49.

Flujo de trabajo de UC Irvine: Baseline Digital Accessioning Workflow for Logical Directories

Mientras el equipo de revisión de los *Guidelines* estaba terminando el proyecto, tuve la oportunidad de recibir a Carolina Quezada Meneses—una becaria que cursaba su maestría en archivos (MLIS) en la Universidad de California, Los Ángeles (UCLA)—para ayudarme a investigar y actualizar nuevos flujos de trabajo en el Departamento de Colecciones Especiales y Archivos en la Universidad de California, Irvine (UCI). El flujo de trabajo que existía se desarrolló en 2013 y dependía del uso de una estación de trabajo singular y del uso de la línea de comando para ejecutar un script complejo para llevar a cabo todas las acciones de procesamiento (escaneo de virus, generación de imágenes de disco, generación de metadatos y empaquetado). Durante muchos años, este flujo de trabajo sirvió como un servicio único para procesar todos los formatos nacidos digitales en nuestro departamento.

Sin embargo, hubo huecos significativos en este flujo de trabajo, incluido el hecho de que esencialmente preparaba materiales nacidos digitales para la *preservación*, pero no abordaba cómo proporcionar *acceso* a esos materiales. El script ejecutado en la línea de comando también se rompía frecuentemente y, debido a la lógica lineal del script, todo el trabajo tenía que detenerse cuando eso pasaba hasta que colegas ajenos al departamento pudieran repararlo. Finalmente, la estación de trabajo estaba físicamente ubicada fuera del departamento y la mayoría de las imágenes de disco fueron creadas por personal ajeno al departamento, lo que generó una serie de retrasos en nuestro trabajo.

Tool/Software Name	Maintained By	Function	Where is it installed?
Windows Defender	Microsoft	Installed on all networked staff computers, managed by Library IT. You	All networked staff

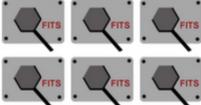
<p>Security Center</p> 		<p>can run a voluntary virus scan by selecting “Run” in the Scan History / Virus & threat protection section.</p>	<p>computers</p>
<p>Directory List & Print</p> 	<p>INFONAUTICS GmbH Switzerland</p>	<p>Free software for Windows. Enables the listing and printing of a directory structure. Separates folder titles into comma-separated values (CSV), which allows for easy repurposing of description. Replaces Karen’s Directory Printer in our workflows.</p>	<p>Tech Services workstations and C816 and C035</p>
<p>Bulk_extractor</p> 	<p>Garfinkel, Simson on GitHub</p>	<p>Free tool available for Linux, OS X, and Windows that scans a disk image, a file, or a directory of files and extracts keyword phrases to make PII searching possible. Offered as part of the BitCurator suite or can be downloaded as a standalone.</p>	<p>Tech Services workstations and C816 in LL568</p>
<p>FITS</p> 	<p>Harvard University</p>	<p>Free tool available for Windows and OS X. The File Information Tool Set (FITS) identifies, validates and extracts technical metadata about each file and its format. It acts as a wrapper for a suite of open source tools included in the toolset. Output from these tools is consolidated into a single XML output file.</p>	<p>Tech Services workstations</p>
<p>Bagger</p> 	<p>Library of Congress</p>	<p>Bagger is the GUI version of Bag-It, an open source tool that packages, verifies and validates archival content. It zips content and creates hash values on all files in the Payload as well as the entire bag. It is offered in the BitCurator suite or can be downloaded as a standalone.</p>	<p>Tech Services workstations and C816 in LL568</p>

Imagen 2. Herramientas abiertas que utilizamos para nuestro flujo de trabajo

Quando Carolina y yo nos propusimos actualizar nuestro flujo de trabajo, fuimos influenciadas por el mismo sentimiento de “prácticas suficientemente buenas” que prevalece en

las recomendaciones de los *Guidelines*. Nuestro objetivo principal era crear flujos de trabajo y emplear herramientas accesibles para todo el personal, con los recursos disponibles. En nuestro personal existe experiencia diversa y no todos colegas se sienten cómodos con utilizar la línea de comando. Con este nuevo flujo de trabajo, pudimos construir un proceso más modular y flexible disponible en todas las estaciones de trabajo del personal. Con este nuevo flujo de trabajo, el personal ya no tiene que depender del uso de la estación única o de otro personal ajeno al departamento, para completar el trabajo; ahora están autorizados para acceder a sus propios materiales, a su propio ritmo.

Este flujo de trabajo describe los pasos básicos de acceso para los materiales de archivo "puramente digitales" localizados en el departamento de colecciones especiales en la Universidad de California de Irvine. Los materiales de archivo "puramente digitales" constituyen un porcentaje creciente de nuestras nuevas adquisiciones. Los directorios lógicos se consideran "puramente digitales," ya que no están vinculados al operador y consisten en una estructura de directorios organizada por carpetas y archivos. Los directorios lógicos se entregan a través de Google Drive, Box, u otros servicios en la nube; transferencias de red, STP o SFTP; USB o hard drives.

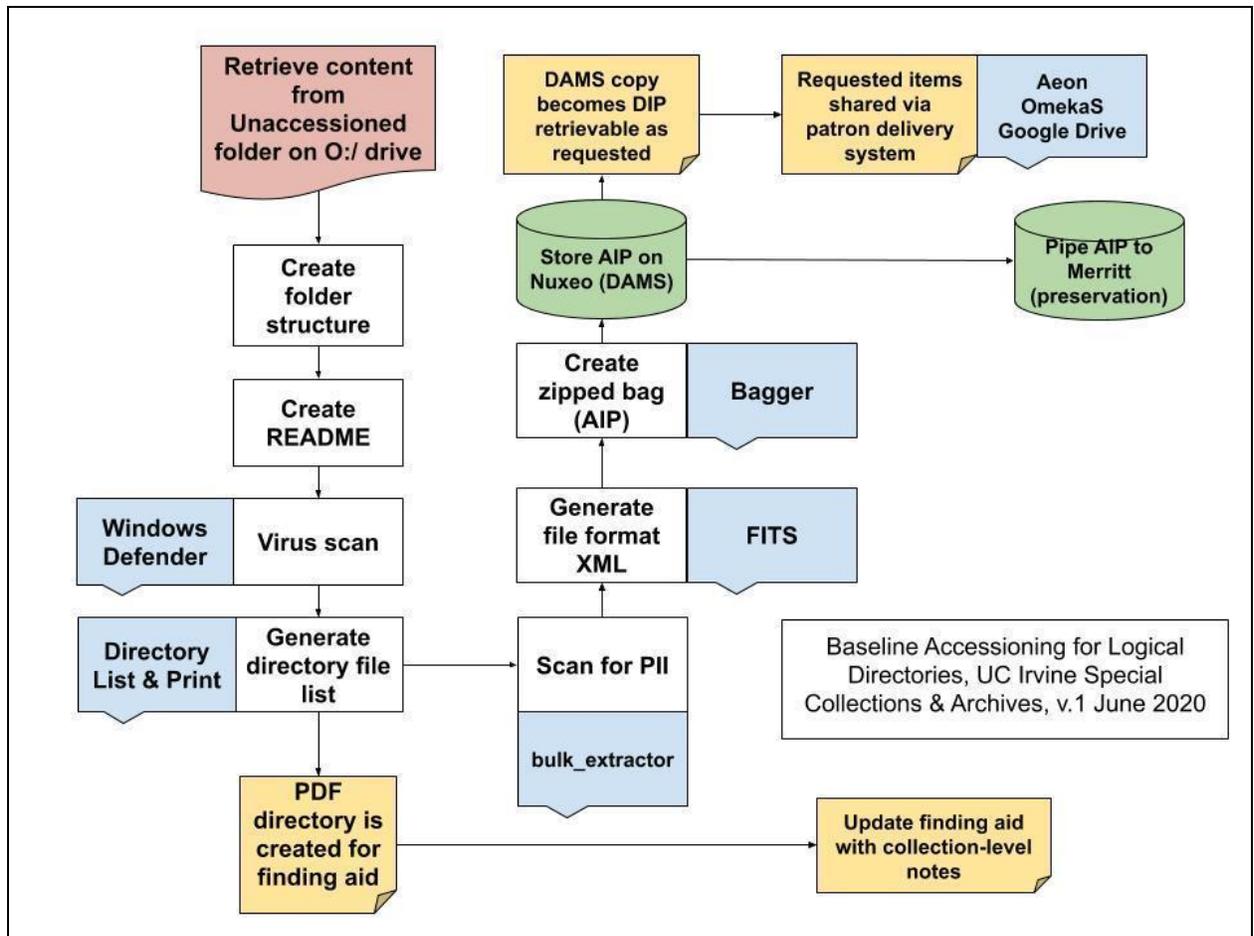


Imagen 3. Flujo de trabajo de UC Irvine: Baseline Digital Accessioning Workflow for Logical Directories

Este flujo de trabajo toma muchas de las recomendaciones en los *Guidelines* y las aplica en varias partes. Por ejemplo, incorpora el uso de herramientas ya instaladas en nuestras estaciones de trabajo como Windows Defender (que el Departamento de Tecnología del campus actualiza regularmente) para la detección de virus. Incluye el uso de herramientas de código abierto (open-source) para generar una lista de archivos de directorio, esta lista de archivos se adjunta al *finding aid* como un PDF que ayuda a exponer la información de nivel de directorio para los usuarios. Ya no creamos imágenes de disco de forma predeterminada (no solo para los directorios lógicos sino que tampoco para otros formatos) y simplemente copiamos o

transferimos estos materiales a nuestros servidores para esperar la accesión. Este flujo de trabajo también aborda el acceso: para los materiales que presentan poco o ningún riesgo conocido de identificación personal, datos confidenciales o sensibles, podemos proporcionar acceso a estos materiales solo con la accesión, y el procesamiento más intensivo está reservado solo para los materiales que presentan un riesgo de confidencialidad o sensibilidad.

En medio del desarrollo de estos flujos de trabajo tuvimos que trabajar desde el hogar debido a la pandemia de COVID-19, y, por lo tanto, aún tenemos que completar la actualización de nuestros flujos de trabajo en lo que respecta al procesamiento de materiales nacidos digitales. El trabajo de procesar archivos digitales depende de software instalado sólo en la oficina. Sin embargo, debido a que este flujo de trabajo de accesión ya está disponible para el personal a través de *Remote Desktop*, este ha sido increíblemente útil para estar al tanto de nuestros materiales entrantes y la acumulación digital mientras trabajamos desde casa.

Casos en que la imagen de disco es apropiado: preservar el arte digital de Christine Tamblin

Aunque la archivística se está distanciando de la práctica de la creación de imagen de disco por defecto, esta acción sigue siendo un paso fundamental e importante en ciertos casos, por ejemplo como el ya mencionado caso para preservar software único. El arte digital en el que el software es parte de la experiencia artística y se necesita preservar es uno de los casos más comunes en donde la imagen de disco es “una poderosa herramienta para encapsular tanto la obra de arte como su ambiente de software para su conservación o documentación.”²² Otro proyecto con el que Carolina y yo colaboramos es la preservación y acceso al usuario de obras de

²² Eddy Colloton, Jonathan Farbowitz, Flaminia Fortunato, y Caroline Gil, “Towards Best Practices In Disk Imaging: A Cross-Institutional Approach” *Electronic Media Review*, Volume Six: 2019-2020. <https://resources.culturalheritage.org/emg-review/volume-6-2019-2020/colloton/>.

arte basadas en software creadas por la artista Christine Tamblyn. Nos enfocamos en dos CD-ROMs formateados para Mac que contenían obras de arte basadas en software, *She Loves It, She Loves It Not: Women and Technology* (1993), y *Mistaken Identities* (1995). Los dos programas son ejemplos de arte digital temprano que apareció en las décadas de 1980 y 1990 cuando artistas utilizaron formas experimentales. Los dos programas son interactivos y dependen de la interacción del usuario usando el ratón para que ciertos textos, sonidos, clips e imágenes se muevan o toquen. Por su alta dependencia a la interactividad, y por estar formateados para un ambiente antiguo de Mac, estos CD-ROMs requerían capturar un ambiente completo para poder utilizar un emulador compatible. Carolina creó dos imágenes de disco con nuestro equipo “Bilbo,” una computadora PowerMac G3 en ambiente de Macintosh OS 9.2.2. (ver imagen 4) y experimentó con varios emuladores para asegurarse de que las imágenes de disco pudieran montarse. Para evitar dificultades para los usuarios en acceder contenido en el formato de la imágenes (.IMG) Carolina decidió hacer *screencasts* grabándose ella interactuando con los programas y subirlos a un sitio web que creó para mostrar los programas en los CD-ROMs como una forma de acceso rápido para el usuario.²³

²³ Carolina Quezada Meneses, Christine Tamblyn’s Interactive CD-ROMs artworks, 2020. <https://sites.google.com/view/tamblynproject/home?authuser=0>.

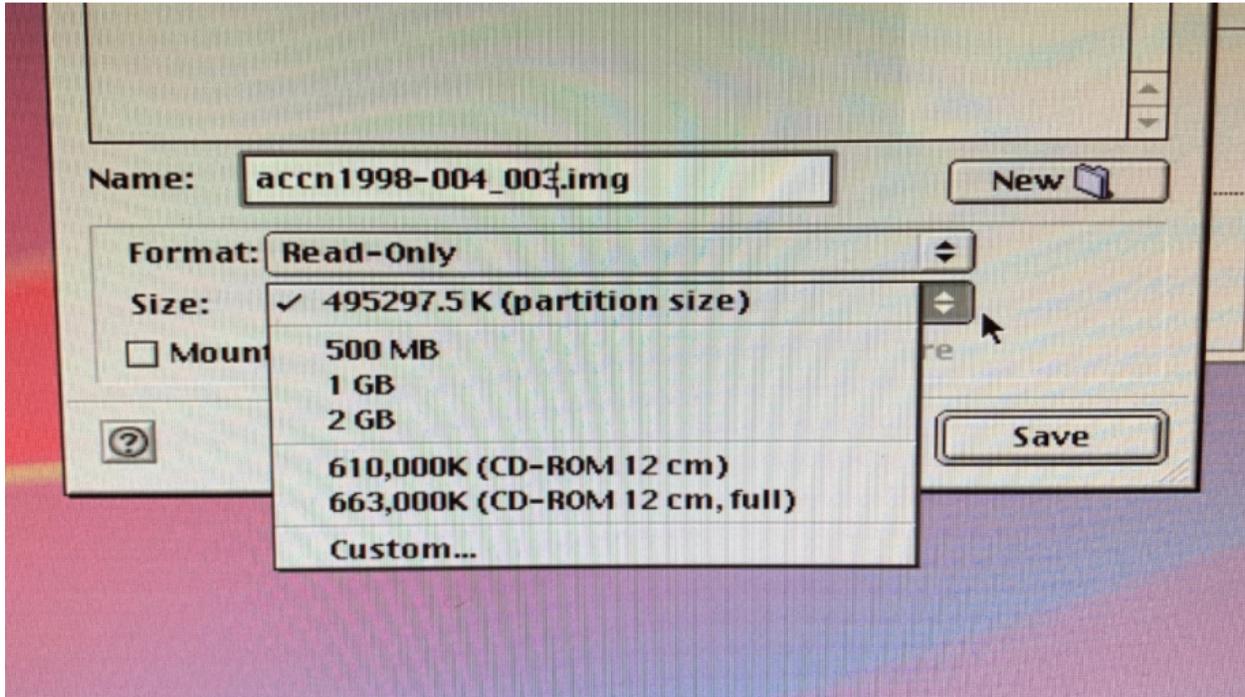


Imagen 4. Captura de pantalla de Bilbo, computadora PowerMac G3 en ambiente de Macintosh OS 9.2.2. para capturar imágenes de disco utilizando DiskCopy 6.3.3. Imagen tomada por Carolina Quezada Meneses.

Conclusión

La mejor práctica de la creación de imágenes de disco por defecto está experimentando una evolución actual donde los archiveros están más conscientes del impacto que esta acción tiene en los recursos laborales, ambientales, y la responsabilidad a la confianza que tenemos que guardar con nuestros donantes, creadores de archivo, y sujetos. La examinación crítica es requerida para evaluar casos en donde sí es verdaderamente necesario y la archivística está actualmente produciendo más recursos para discutir.

Los archiveros han tenido que adaptar su trabajo a un ambiente doméstico, debido al distanciamiento físico resultado del brote y proliferación de COVID-19. En tiempos previos a la pandemia era casi imposible imaginar y legitimar este arreglo para muchos que trabajan con

archivos y colecciones digitales; aunque la mayor porción del trabajo necesario para preservar y preparar acceso a colecciones nacidas digitales se puede hacer desde cualquier sitio con las herramientas adecuadas. Tuvo que hacerse sin aviso o previsión. Será interesante ver cómo los flujos de trabajo para archivos nacidos digitales seguirán viéndose afectados por el trabajo a distancia. Espero que estos dos eventos en la archivística nos lleven hacia flujos de trabajo más ligeros, modulares, sostenibles y responsables.

Bibliografía

- Arroyo-Ramírez, Elvia. “Invisible Defaults and Perceived Limitations: Processing the Juan Gelman Files,” (Preservation and Archiving Special Interest Group (PASIG), New York, October 28, 2016),
<https://medium.com/on-archivy/invisible-defaults-and-perceived-limitations-processing-the-juan-gelman-files-4187fdd36759>.
- Arroyo-Ramírez, Elvia, Kate Dundon, and Shira Peltzman. “Holistic approaches to born-digital appraisal and accessioning—revising the UC Guidelines for Efficient Archival Processing,” presented at *OCLC Works in Progress* webinar, March 16, 2021,
<https://escholarship.org/uc/item/7k45c35j>.
- Colloton, Eddy, Jonathan Farbowitz, Flaminia Fortunato, and Caroline Gil. “Towards Best Practices In Disk Imaging: A Cross-Institutional Approach” *Electronic Media Review*, Volume Six: 2019-2020.
<https://resources.culturalheritage.org/emg-review/volume-6-2019-2020/colloton/>.
- Disk Imaging Decision Factors. Digital Archival Transfer, Ingest, and Packaging Group (DANNNG). Version 0.3
https://docs.google.com/document/d/1g45L8fH2d_tYO0ME1gm-m1obM-Q5fIIPJVMYK6Kr5e0/edit#heading=h.6gwzqgcmnlyz.
- Dundon, Kate, Laurel McPhee, Elvia Arroyo-Ramírez, Jolene Beiser, Courtney Dean, Audra Eagle Yun, Jasmine Jones, Zachary Liebhaber, Charlie Macquarie, Lara Michaels, Shira Peltzman, Liz Phillips. (2020). “Guidelines for Efficient Archival Processing in the University of California Libraries” (Version 4). *UC Office of the President: University of California Systemwide Libraries*. Retrieved from
<https://escholarship.org/uc/item/4b81g01z>.
- Kirschenbaum, Matthew. *Mechanisms: New Media and the Forensic Imagination* (Cambridge, MA: MIT Press, 2008).
- Lassere, Monique and Jess M. Whyte. “Balancing Care and Authenticity in Digital Collections: A Radical Empathy Approach to Working with Disk Images,” in “Radical Empathy in Archival Practice,” eds. Elvia Arroyo-Ramirez, Jasmine Jones, Shannon O’Neill, and Holly Smith. Special issue, *Journal of Critical Library and Information Studies* 3 (pre-print). Retrieved from
<https://journals.litwinbooks.com/index.php/jclis/article/view/125/80>.
- Pendergrass, Keith L., Walker Sampson, Tessa Walsh, and Laura Alagna. (2019). “Toward Environmentally Sustainable Digital Preservation.” *The American Archivist* 82 (1): 165–206. <https://doi.org/10.17723/0360-9081-82.1.165>.

Quezada Meneses, Carolina. (2020). Christine Tamblyn's Interactive CD-ROMs artworks.
<https://sites.google.com/view/tamblynproject/home?authuser=0>.

Woods, Kam, Cal Lee, y Simson Garfinkel. (2011). "Extending Digital Repository Architectures to Support Disk Image Preservation and Access," en *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries* (New York: Association of Computing Machinery). <https://doi.org/10.1145/1998076.1998088>.