

A stylized yellow circuit board graphic is positioned on the left side of the cover, extending towards the center. It features various components like resistors, capacitors, and traces, all rendered in a clean, minimalist style. The background is a solid, vibrant green.

La teoría y el oficio de la **preservación digital**

Trevor Owens

Coordinación

Isabel Galina Russell

Universidad Nacional Autónoma de México

La teoría y el oficio de la preservación digital

Trevor Owens

La teoría y el oficio de la preservación digital

Trevor Owens

Coordinación

Isabel Galina Russell

Traducción

Pedro Ángeles Jiménez

Fernando Cruz Quintana

Jimena Escobar Sotomayor

Isabel Galina Russell

Maribel González González

David Alonso Leija Román

Norma Aída Manzanera Silva

Ana Cecilia Medina Arias

Betsabé Miramontes Vidal

Jo Ana Morfin

Ana Yuri Ramírez Molina

Lorena Ramírez López

José Antonio Salazar Carmona

Pamela Vizner Oyarce



Ciudad de México, 2023

Biblioteca Nacional de México. Catalogación en Publicación (CIP).

Nombres: Owens, Trevor, autor. | Galina Russell, Isabel, coordinador. | Ángeles Jiménez, Pedro, traductor. | Universidad Nacional Autónoma de México. Instituto de Investigaciones Bibliográficas, editor.

Título: La teoría y el oficio de la preservación digital / Trevor Owens ; coordinación Isabel Galina Russell ; traductor Pedro Ángeles Jiménez [y otros].

Descripción: Primera edición. | Ciudad de México : Universidad Nacional Autónoma de México, Instituto de Investigaciones Bibliográficas, Biblioteca Nacional, Hemeroteca Nacional, 2023. | Serie instrumenta bibliographica ; 4

Identificadores: 978-607-30-7827-6

Temas: Preservación digital. | Materiales bibliotecarios -- Conservación y restauración.

Clasificación: 025.84 cdd22

No. de Registro BNM: 737615

Diseño de forros: Hilda Maldonado

Primera edición en español (impresa): 2022

Primera edición digital en español (PDF): 2023

D. R. © 2022 Universidad Nacional Autónoma de México

Instituto de Investigaciones Bibliográficas

Biblioteca Nacional / Hemeroteca Nacional

Centro Cultural Universitario, Ciudad Universitaria

Alcaldía Coyoacán, C. P. 04510, Ciudad de México

Tel. (55) 5622 6811 / www.iib.unam.mx

ISBN (impreso): 978-607-30-7031-7

ISBN (PDF): 978-607-30-7827-6

ISBN (ePub): 978-607-30-7826-9

Título original: *The theory and craft of digital preservation*

ISBN (rústica): 9781421426976

Publicado mediante acuerdo con Johns Hopkins University Press

Primera edición en inglés

© 2018 Johns Hopkins University Press

Todos los derechos reservados.

Johns Hopkins University Press

2715 North Charles Street

Baltimore, Maryland 21218-4363



La teoría y el oficio de la preservación digital por Universidad Nacional Autónoma de México se distribuye bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional. Basada en una obra en <http://www.iib.unam.mx>.

Hecho en México / Made in Mexico

Contenido

Serie Instrumenta Bibliographica	10
Presentación de la edición	11

La teoría y el oficio de la preservación digital	17
---	----

Agradecimientos	19
-----------------	----

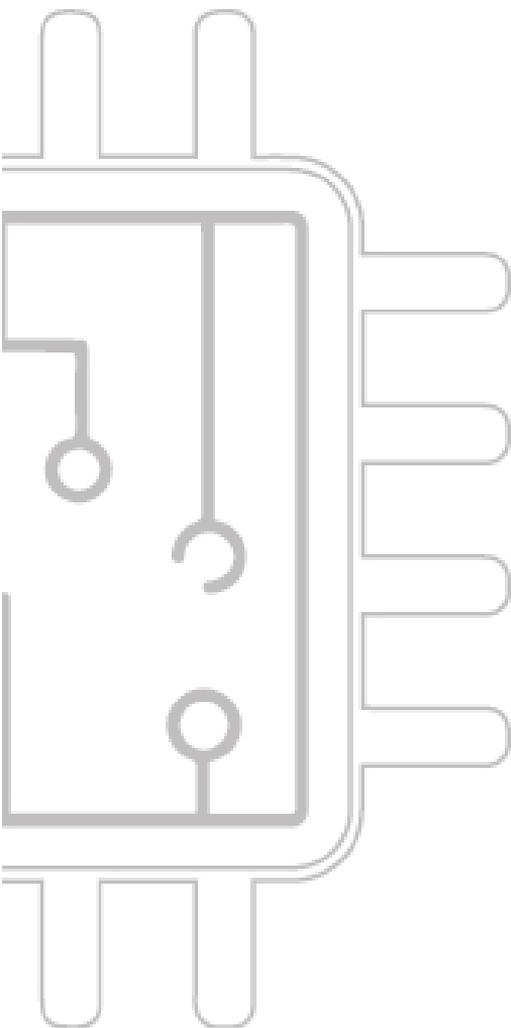
Introducción	
Más allá del furor y la ansiedad digital	23

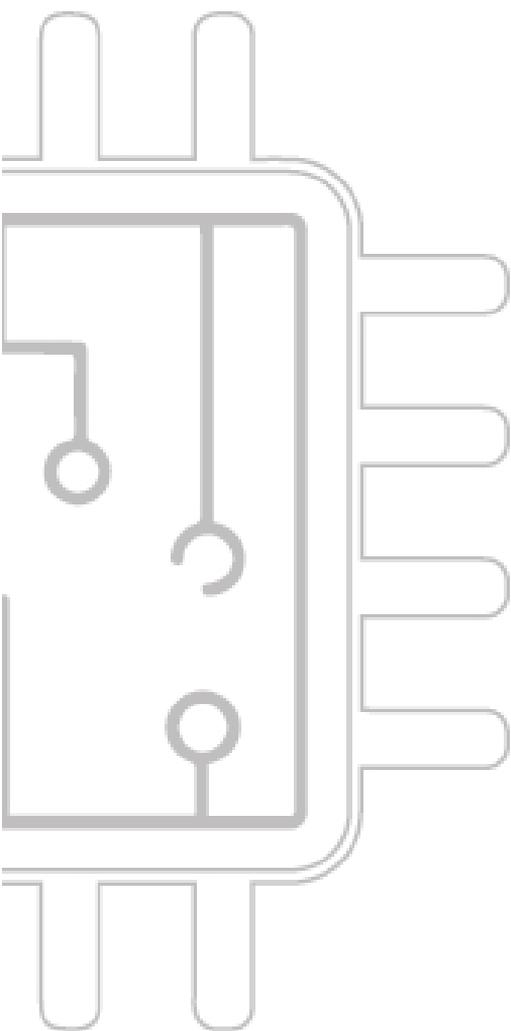
Capítulo 1	
Las tradiciones divergentes de la preservación	33

Capítulo 2	
La importancia de entender los objetos digitales	55

Capítulo 3	
Los desafíos y las oportunidades de la preservación digital	75

Capítulo 4	
El oficio de la preservación digital	91





Capítulo 5	
El propósito de la preservación y el desarrollo de colecciones	101
Capítulo 6	
La gestión de copias y formatos	121
Capítulo 7	
La organización y la descripción de los objetos digitales	145
Capítulo 8	
La habilitación de los usos y los accesos multimodales	175
Conclusiones	
Herramientas hacia el futuro	201
Bibliografía	215

SERIE INSTRUMENTA BIBLIOGRAPHICA

La histórica serie Instrumenta Bibliographica dio pie a la publicación, bajo la responsabilidad de distinguidos investigadores del Instituto de Investigaciones Bibliográficas, de libros pensados como herramientas para la investigación bibliográfica y hemerográfica sobre el patrimonio nacional. En 1973 salieron a la luz los dos primeros títulos: *Manual de metodología y técnica bibliográficas*, de Gloria Escamilla, e *Índice de nombres latinos de ciudades con imprenta (1448-1825)*, de José Ignacio Mantecón Navasal. Entre 1976 y 1978, la maestra Escamilla publicó en dos volúmenes la traducción de algunos capítulos de las *Reglas de catalogación angloamericanas* y en 1979 preparó el manual *Interpretación catalográfica de los libros*. En 1985 María del Carmen Ruiz Castañeda y Sergio Márquez formaron el *Catálogo de seudónimos, anagramas, iniciales y otros alias usados por escritores mexicanos y extranjeros* y en 1990 editaron sus *Correcciones*. En total, se trata de siete títulos que se convirtieron en obras de consulta imprescindibles para la investigación bibliográfica y bibliotecológica.

Al cumplirse 30 años de la primera edición del último volumen de la serie inicial, y ante los cambios vertiginosos en el campo de la investigación humanística debidos, en gran parte, a las importantes innovaciones tecnológicas y su impacto en la configuración de la cultura escrita, así como a los múltiples senderos que se vislumbran en su aprehensión, la Biblioteca Nacional de México ha decidido retomar, con nuevo impulso, la publicación de una serie de obras destinada a ofrecer información precisa y pautas metodológicas apuntaladas por estudios y buenas prácticas en torno al mundo del libro, las bibliotecas y todos aquellos temas vinculados por un mismo denominador común: el conocimiento y la preservación del patrimonio documental.

Presentación de la edición

Tuve la oportunidad de escuchar a Trevor Owens mientras presentaba sus axiomas para la preservación digital durante un encuentro en la Escuela Nacional de Conservación, Restauración y Museografía a finales del 2019. Envié un tuit: “16 axiomas por Trevor Owens para preservación digital. Claros, sensatos, prácticos @tjowens en el #encrym. Habría que encontrar la forma de traducir su libro al español”. Pocas horas después Trevor respondió, y así nació la idea de publicar su libro, *The Theory and Craft of Digital Preservation*, en español.



Intercambio de tuits,
diciembre de 2019.



Trevor Owens nos facilitó el contacto con la editorial Johns Hopkins University Press, quienes nos apoyaron en adquirir el permiso para realizar la traducción de la obra. En el Instituto de Investigaciones Bibliográficas de la Universidad Nacional Autónoma de México, donde trabajo como investigadora, mostraron interés por publicarlo en la serie *Instrumenta Bibliographica* de la Biblioteca Nacional. Sin embargo, no contábamos con el presupuesto para traducir el libro.

Traducir es un arte y un oficio. En la academia internacional, donde el inglés es el idioma prevalente para las comunicaciones y las publicaciones de la comunidad global, existen barreras importantes para aquellos que quieren acceder al conocimiento y ser parte de las discusiones en este ámbito. El monolingüismo académico es un problema que frecuentemente no se reconoce. Por el contrario, se asume que las personas que quieren participar deben aprender inglés para ser incluidas, por lo que, en esta cerrazón, se pierden voces y experiencias valiosas. Existen formas de remediar esta situación, y la traducción es una de ellas. Lamentablemente, suele ser un oficio mal pagado e invisibilizado. En los presupuestos académicos rara vez encontramos partidas o asignación de recursos para este tipo de trabajos, pese a que uno esperaría que el acceso al conocimiento publicado en todos los idiomas fuera una prioridad para nosotros. Hablamos de la importancia de la multiculturalidad, pero las barreras lingüísticas la impiden.

En este libro, Trevor Owens, refiriéndose a la catalogación de materiales digitales, nos advierte sobre el peligro de no hacer nada, en espera de hacerlo todo perfecto. Busquemos entonces otras maneras de manejar las cosas, soluciones creativas que nos permitan avanzar de diferentes formas, pero logrando el mismo objetivo. En el campo de las humanidades digitales, éste ha sido un tema ampliamente discutido, tanto en términos teóricos como prácticos. Un ejemplo de una solución útil para estas dificultades es *Translation Toolkit*,¹ donde se ofrecen opciones prácticas y sencillas que aprovechan las plataformas digitales para facilitar la comunicación académica en congresos.

En 2018 se publicó un artículo en la revista *Wired* sobre *Torn Apart*, un proyecto digital que surgió como respuesta a la política “tolerancia cero”, implementada por el gobierno estadounidense y que significó la inhumana y terrible separación forzada de miles de niños y sus familias. Consideramos que ese

¹ Para más información al respecto, ver el trabajo de Gimena del Río, Alex Gil, Daniel O'Donnell y Élika Ortega, “About”, *Translation Toolkit*, acceso el 28 de abril de 2021, <https://go-dh.github.io/translation-toolkit/>.



texto podía ser de interés no sólo para los humanistas digitales que trabajan en México, sino también para el público en general. Tanto los autores como los editores de la revista apoyaron nuestra iniciativa y nos dieron los permisos para realizarla. Si bien muchos de nosotros estábamos interesados y comprometidos, no necesariamente contábamos con el tiempo, los recursos o el conocimiento para traducir todo el artículo. De ahí, inspirados en la idea de Toolkit, realizamos un experimento de traducción colaborativa en el que conformamos un grupo de personas interesadas en trabajar en este proyecto. Dividimos el artículo en párrafos, de tal suerte que cada quien tradujo dos o tres. Al último, se revisó el texto completo para dar estilo y congruencia a la versión final.²

Pensé que podíamos emplear este método para hacer la traducción del libro. En el 2017 participé en la fundación del Grupo de Preservación Digital (GPD), colectivo multidisciplinario e interinstitucional que busca promover la investigación, experimentación y formación de recursos en torno a la problemática de la preservación digital, en particular, de materiales digitales patrimoniales.³ El GPD es, sobre todo, una comunidad de personas interesadas en esta temática y una excelente red de contactos y conocimiento. Hice un llamado a través de este grupo y la respuesta fue excelente, a tal grado que contamos con 14 colaboradores.

Necesitábamos una plataforma que nos permitiera hacer la traducción. Para el proyecto anterior, que era mucho más corto, habíamos utilizado un documento en Google Drive, pero me pareció necesario encontrar otra opción debido a la extensión del texto a traducir. Tenía algo de familiaridad con FromThePage,⁴ un sistema para la transcripción colaborativa de documentos, y pensé que podíamos adaptarlo a nuestras necesidades. Si bien se veía útil, tenía un costo. Agradezco mucho las gestiones de Albert Palacios, y de Ben y Sara Brumfield, los desarrolladores de la plataforma, quienes generosamente nos permitieron usarla a cambio de nuestra retroalimentación después de experimentar con ella como herramienta de traducción, ya que ellos mismos están interesados en sistemas multilingües.

² Isabel Galina, “‘ICE está en todas partes’: el uso de las ciencias de la información para mapear la crisis de las familias separadas”, trad. de Isabel Galina et al., Red de humanidades digitales (blog), 19 de septiembre de 2018, <http://humanidadesdigitales.net/blog/2018/09/19/ice-todas-partes-uso-ciencias-informacionmapear-crisis-familias-separadas/>

³ “Integrantes”, Grupo de Preservación Digital, acceso el 28 de abril de 2021, <https://preservaciondigital.iib.unam.mx/txt/integrantes.html>.

⁴ FromThePage, acceso el 28 de abril de 2021, <https://fromthepage.com>.



Realizamos la traducción a partir de la versión de acceso abierto, el *preprint*.⁵ Ingresamos el PDF a la plataforma y los traductores seleccionaron el o los capítulos que les interesaba traducir. Cada capítulo debía tener dos traducciones. FromThePage permite acceder a dos ventanas contiguas, lo que hace posible visualizar el original y la traducción, o ambas traducciones, a la par. Más adelante, como coordinadora, revisé las dos traducciones y las utilicé para crear una sola. Yo hablo dos idiomas, pero eso no significa que sea traductora, así que consideré que tanto el original como el formato de publicación ameritaban una revisión profesional, la cual fue realizada por Anna Cristina Casas, quien le dio uniformidad y calidad a todo el texto. A su vez, Fernanda Baroco, encargada de la corrección de estilo por parte del Departamento Editorial, revisó la traducción nuevamente, cotejando el *preprint* con la versión final publicada. Sin duda, esta edición es resultado de un esfuerzo colectivo. En el camino hemos aprendido mucho sobre cómo trabajar de manera colaborativa: la importancia del orden, del diálogo, de los consensos, de documentar los procesos y también de cómo las plataformas digitales nos pueden ayudar a llevar a cabo estas tareas de forma distribuida.

Traducir el libro al español es un intento para que este valioso texto sea accesible a un mayor número de personas. Sin embargo, casi todos los campos relacionados con las tecnologías digitales cambian rápidamente y se generan nuevos términos con rapidez, casi siempre en inglés, desde una lógica neoliberal, como bien nos advierte Owens. Estamos entonces, con este libro, fijando expresiones, creando nomenclaturas, y eso es también una postura. Elegir dejar una palabra en inglés o generar otra en español no es una tarea trivial. Sabemos que estamos participando en el establecimiento de neologismos. Al realizar la traducción, surgieron debates y conversaciones importantes en torno a cómo consignaríamos ciertas voces. Por ejemplo, *remediation* aquí aparece como “remediación” y no como “remediatización”. Aunque en el diccionario la acepción que estamos utilizando aún no ha sido incorporada, sí encontramos ejemplos de su uso en publicaciones académicas.⁶ El término en inglés *fixity* se traduce en ocasiones como “integridad”, pero en otras como “fijeza”, dependiendo de la forma en que es empleado. La traducción es también una

5 Trevor Owens, “The Theory and Craft of Digital Preservation” (manuscrito inédito, junio de 2017), <https://osf.io/preprints/lissa/5cpjt>.

6 Por ejemplo, ver la traducción de Eva Aladro del artículo “Inmediatez, hipermediación, remediación”, *Cuadernos de Información y Comunicación* 16 (2011): 29-57, <https://revistas.ucm.es/index.php/CIYC/article/view/36986>.



interpretación y una remediación (en nuestro nuevo sentido), y es un acto subjetivo. Además de brindar a la comunidad hispanoparlante la posibilidad de leer este fantástico libro de Trevor, esperamos también haber contribuido al campo de la preservación digital.

No puedo cerrar esta presentación sin una reflexión sobre el momento histórico en el cual se enmarcó este proyecto. Trevor y yo intercambiamos comunicaciones entre enero y marzo del 2020. Elaboré un cronograma para la traducción de este libro, en el cual consideré unos meses para su desarrollo, sin saber que al poco tiempo mi mundo y el de muchos otros estaba por cambiar con la declaración de emergencia sanitaria por covid-19. Esta obra se demoró mucho más de lo esperado.

Considero, sin embargo, que el momento para su publicación es más acertado que nunca. La pandemia exacerbó nuestro uso de los medios digitales de una forma sin precedentes y, al trasladar toda nuestra actividad a estas plataformas, más que nunca, creamos materiales digitales que forman parte de nuestra cotidianidad, nuestra herencia y nuestro patrimonio. Las inquietudes que menciona Trevor sobre la importancia de preservar lo que es importante para nosotros, para nuestras comunidades, se agudizan cuando pensamos en más de dos años de una producción digital sin parangón. Esperamos que la traducción de este libro sea una aportación significativa para auxiliar a las personas que se preocupan y se ocupan de esta importante tarea.

Isabel Galina Russell
diciembre de 2022

La teoría y el oficio de la preservación digital

Trevor Owens

Agradecimientos

He pasado un año trabajando en este libro, pero representa la culminación de casi una década de intentar encontrar mi propio sentido de lo que es la preservación digital. Es por eso que tengo muchas personas a quienes agradecer. Las fortalezas de este libro vienen de la comunidad internacional de preservación digital en la que fui acogido. Las debilidades del libro son sólo mías.

Mi primer acercamiento a la preservación digital fue durante el tiempo que pasé en el Roy Rosenzweig Center for History and New Media (RRCHNM). Antes de fallecer, Roy tuvo un impacto importante y duradero en aquellos que tuvimos la fortuna de trabajar con él. Tengo el anhelo de que la compasión, la dedicación y el pragmatismo que Roy ejerció en sus labores diarias en el RRCHNM se transmitan a través de mi trabajo. Mi conocimiento y apreciación por los temas de la historia digital y la preservación digital surgieron a partir de cuatro años de discusión y colaboración con mis colegas de ese lugar, Dan Cohen, Josh Greenberg, Sean Takats, Tom Scheinfeldt, Sharon Leon, Sheila Brennan, Dave Lester, Jeremy Boggs, Jim Safley, Kari Kraus, Connie Moon Sehat, Miles Kelly, Mindy Lawrence, Jon Lesser, Kris Kelly, Ken Albers, Faolan Cheslack-Postava, John Flatness, Dan Stillman y Christopher Hamner. Fue en el curso de posgrado de historia digital de Dan que escuché por primera vez cosas como “descomposición de bits”, “degradación elegante” y “preservación digital” de manera formal. Cuando Sharon no pudo asistir a la reunión del National Digital Information Infrastructure and Preservation Program (NDIIPP) de la Library of Congress, yo fui en su lugar y tuve la oportunidad de conocer a muchos de mis futuros colegas. Josh fue muy generoso al ofrecerse a realizar comentarios sustantivos para mi tesis de licenciatura,



siempre estuvo dispuesto a charlar acerca de estudios de ciencia y tecnología, y, además, fue mi referencia cuando solicité trabajo en el NDIIPP.

Cuando me integré al equipo del NDIIPP en el 2010, no tenía idea de lo afortunado que era. Tuve la oportunidad de unirme a un equipo constituido por algunas de las personas más brillantes en el mundo que trabajan en la preservación digital. La mejor parte es que todos eran sorprendentemente accesibles. En mi primer día de trabajo, Dan Chudnov (cuyo artículo “Emperor’s New Repository” continúa siendo de los mejores textos sobre la preservación digital) me comentó que “puedes amar a la Library of Congress, pero nunca te amarás de vuelta”. Es cierto. Las instituciones masivas como la Library of Congress pueden sentirse como burocracias frías. Sin embargo, esos retos pueden crear lazos y amistades que perduran. Los mejores tesoros de la Library of Congress se encuentran en su personal. Mis colegas del NDIIPP, y del resto de la institución, me enseñaron qué es la preservación digital y cómo se lleva a cabo realmente. Por ello, debo un agradecimiento considerable a Martha Anderson, Abbey Potter, Michelle Gallinger, Jane Mandelbaum, Dan Chudnov, Butch Lazorchek, Erin Engle, Leslie Johnston, Jefferson Bailey, Abbie Grotke, Ed Summers, Nicholas Taylor, Chris Thatcher, Niki Saylor, Chris Adams, Thomas Padilla, Emily Reynolds, Gloria Gonzalez, Jimi Jones, Bertram Lyons, Kate Murray, Caroline Arms, Carl Fleischhauer, Caitlin Hammer, Andrew Cassidy-Amstutz, Kathleen O’Neill, Meg McAleer, Moryma Aydelott, Beth Dulabahn, Kate Zwaard, David Brunton y muchos más.

Mientras trabajaba para el NDIIPP, tuve la oportunidad de desempeñarme como el primer codirector del grupo de trabajo de preservación digital, denominado National Digital Stewardship Alliance (NDSA). Durante nuestras llamadas por conferencias mensuales, en las que discutíamos qué funcionaba y qué no, se consolidaron muchas de las ideas para este libro. Aunque todos los involucrados en las reuniones y los eventos relacionados con la NDSA tuvieron un impacto en mi forma de pensar, me gustaría mencionar en particular a Micah Altmen, Andrea Goethals, Andy Johnson, Lori Emerson, Jason Eppink, Amanda Brennan, Euan Cochran, Megan Phillips, Karen Carani, Dragan Espenschied, Cal Lee, David Rosenthal, Vicky Reich, Katherine Skinner, Nick Krabbenhoft, Robin Ruggaber, Cory Snavely, Peter Krogh, Micheal Edson, Mike Giarlo, David Pearson, Bethany Nowvickie, Robert Horton, Chris Lacinak, Kara Van Malssen, Abby Smith Rumsey, Howard Besser, Don Watters, Doug Reside, Ian Bogost, Henry Lowood, Jason Scott, Richard Pierce-Moses, Mark Matienzo, Maureen Callahan, Jaime Schumacher, Stacey Erdman, Seb Chan y Aaron Straup Cope.



También me gustaría agradecer a las personas que se tomaron el tiempo para comentar los borradores del libro que compartí en mi blog, Bertram Lyons, Alan Liu, Chris Adams, Karl-Rainer Blumenthal, Porter Olsen, Matthew Kirschenbaum, Jon Ippolito, Thomas Padilla, Jessica Tieman, Glynn Edwards, Andrew Jackson, Euan Cochrane, Shira Peltzman, Clarissa Ceglie, Annie Johnson y Steven Lubar. Este libro se fortaleció significativamente gracias a sus aportaciones.

A través de todos estos lugares y espacios he tenido la fortuna de tener un grupo de mentores que me han enseñado mucho sobre cómo trabajan las instituciones culturales: Roy Rosenzweig, Dan Cohen, Josh Greenberg, Martha Anderson, Carl Fleischhauer, Kim Sheridan, Richard Staley y Maura Marx.

Debo también agradecer a mi editor, Matt McAdam. Al principio discutimos la idea de este libro mientras tomábamos café en la Library of Congress en 2014. Me fui de ahí y terminé escribiendo otro libro, uno totalmente diferente, pero todo ese tiempo seguía pensando sobre lo que podría ser este libro. Matt y todo el equipo de la Johns Hopkins University Press han sido espléndidos para trabajar a lo largo de este proyecto. En particular, el libro se ha beneficiado de manera significativa por las ediciones de la coordinadora editorial Juliana McCarthy y Laura Dewey.

Por último, y sobre todo, me gustaría agradecer y reconocer a Marjee Chmiel. Mi esposa. Mi compañera constante. Fuiste tú quien leyó los primeros borradores incoherentes de este libro y me ayudaste a mejorarlos. Hiciste lo mismo para mis cartas de presentación y currículum para cada uno de los trabajos que he mencionado. Me trajiste a DC desde Wisconsin. Estás tan presente en mi vida y mi trabajo que es imposible articular cosas específicas que te pueda atribuir. Lo que es seguro es que no tengo idea de quién sería o qué estaría haciendo fuera de esta vida que tenemos juntos. Escribimos juntos nuestras tesis. Nos turnamos escribiendo y haciendo investigación mientras el otro cocinaba, lavaba los trastes o paseaba a nuestros perros. Tus logros me inspiran. Tu perspectiva profundiza mis ideas. Tu consideración y compasión me animan a trabajar para ser una mejor persona.

Introducción

Más allá del furor y la ansiedad digital

A principios de la década del 2000, una cumbre sobre preservación digital en la Library of Congress reunió a los líderes de la industria y al sector dedicado al patrimonio cultural para trabajar en la resolución de los crecientes desafíos para garantizar el acceso, a largo plazo, a la información digital. Un participante de una compañía de tecnología propuso algo como “¿por qué no sólo juntamos todo eso y lo disparamos al espacio?”. En este caso, “eso” se refería a todo el contenido digital históricamente significativo. A muchos de los participantes les dio risa, pero no lo decía en broma. Muchas personas han buscado, y continúan buscando, de manera similar (aunque no tan literalmente) resolverlo “de un disparo”, una sola solución técnica al problema del acceso duradero a la información digital.

Casi dos décadas después, nos encontramos en medio de las mismas historias y soluciones imaginadas que hemos escuchado durante (al menos) los últimos 20 años. Existe entre la gente la creencia (y preocupación) de que si algo está en Internet, seguirá presente para siempre. Al mismo tiempo, la advertencia sobre una inminente “era oscura”, en la que los registros del pasado reciente se perderán por completo o serán inaccesibles, aparece con regularidad en la prensa masiva. Es como si la información digital durara para siempre, pero, al mismo tiempo, pudiera desaparecer dramáticamente en cualquier momento. Los ciclos del furor de la tecnología digital, combinados con una falta de comprensión básica sobre los medios digitales, nos dejan mal equipados para sortear el furor y la ansiedad. Sin embargo, he descubierto que, cuando le digo a la gente que trabajo en la preservación digital y le explico lo que quiero decir con ello, la mayoría responde algo así como “¡Vaya!, ¡ni siquiera había pensado en eso!”.



Muchos ejecutivos, formuladores de políticas y administradores novatos en el tema creen que el mundo necesita de alguien que diseñe un supersistema capaz de “resolver” el problema de la preservación digital. La sabiduría acumulada de los profesionales de este campo en bibliotecas, archivos, museos y otras instituciones de memoria cultural que han estado haciendo este trabajo durante medio siglo sugiere que es un sueño ilusorio. Para ellos, la idea persistente del supersistema tecnológico que “soluciona el problema” es una distracción que no vale la pena perseguir. También es algo que desvía recursos de quienes hacen el trabajo en realidad. Garantizar el acceso a la información digital a largo plazo no es un problema que se resolverá con una sola herramienta. Más bien, es un campo complejo con un conjunto significativo de dimensiones éticas. Es una vocación. Sólo es posible mediante el uso constante de los recursos de nuestras instituciones culturales. Este libro pretende ser un punto de entrada a la teoría y el oficio de la preservación digital tal y como ha surgido en la práctica.

El propósito de este libro es ofrecer un camino para ir más allá del furor y la ansiedad de “lo digital” y así establecer una línea base para la práctica. Para hacer esto, primero hay que desglosar lo que entendemos por “preservación”. Establecer un conocimiento básico de la naturaleza de los medios digitales y de la información digital es fundamental. Con ello a la mano, cualquiera podrá hacer avances significativos y prácticos para mitigar los riesgos más apremiantes de la pérdida digital.

Como guía e introducción, este texto es una síntesis de extensas lecturas, investigaciones, escritos y charlas dedicadas al tema de la preservación digital. Se basa en mi trabajo en la Library of Congress y, antes de eso, en proyectos de humanidades digitales llevados a cabo en el RRCHNM de la George Mason University. La primera sección del libro sintetiza la investigación sobre la historia de la preservación en una variedad de áreas (archivos, manuscritos, audio, etc.) y establece un diálogo entre la historia y los trabajos más novedosos sobre medios, estudios de plataformas y arqueología de medios. Los capítulos posteriores se construyen a partir de este marco teórico como base para un proceso iterativo para la práctica de la preservación digital.

Este libro sirve tanto como una introducción básica a los problemas y prácticas de la preservación digital como un marco teórico para abordarla deliberada e intencionalmente en tanto que es un campo con múltiples vertientes. La audiencia prevista son profesionales de bibliotecas, archivos y museos actuales y emergentes, así como académicos e investigadores que interactúan con ellos. Espero que también tenga una utilidad más amplia para



cualquier persona interesada en comenzar a practicar el oficio de la preservación digital en cualquier otro campo. Si bien los problemas son complejos, he hecho todo lo posible para que el libro pueda ser útil tanto para alguien que trabaja en una institución grande como para un trabajador solitario de una pequeña comunidad histórica o de un archivo.

Si bien el libro tiene una inclinación práctica, no es un manual, pues se volvería obsoleto rápidamente. Tampoco es un conjunto de instrucciones paso a paso, más bien, está pensado como un punto de referencia para desarrollar y perfeccionar el propio oficio de un profesional de la preservación digital. Vale la pena recordar que lo opuesto a lo práctico no es lo teórico, es lo impráctico.¹ Así, las buenas prácticas se desarrollan y definen mejor cuando establecen un diálogo con la teoría, lo que he intentado lograr en este libro. Profundizaré en una variedad de estudios académicos en teoría de medios, así como en teoría de archivos y de biblioteconomía, lo que es un esfuerzo por anclar el oficio de la preservación digital en el entendimiento de las tradiciones de la preservación y la propia naturaleza de los medios y los objetos digitales.

Con eso en mente, usaré la siguiente parte de la introducción para explicar un conjunto de principios que he desarrollado durante el tiempo que llevo trabajando en este campo y que considero que ayudarán a aclarar algunas de las suposiciones problemáticas que surgen en la labor de la preservación digital. No es una guía para el libro, ni es lo que voy a defender. Se trata más bien de un punto de referencia para entender de dónde vengo y a dónde pienso que necesitamos ir.

Dieciséis axiomas rectores de la preservación digital

Como punto de partida, he condensado una guía de 16 axiomas. Estoy consciente de que esto puede sonar un poco pretencioso, pero es la palabra correcta para lo que son. Creo que estos axiomas son puntos que deberían servir como base para el trabajo de la preservación digital. También son una forma útil para definir exactamente qué es y qué no es la preservación digital. Algunos son suposiciones no declaradas que apoyan las perspectivas ortodoxas de la preservación digital; otros no concuerdan con esa ortodoxia. Estos axiomas son elementos para ir desarrollando mientras se avanza en la lectura. Algunos también son puntos que defenderé y demostraré a lo largo del libro. Todos son centrales para el oficio de la preservación digital, desde mi punto de vista.

¹ Terry Cook, “‘The Imperative of Challenging Absolutes’, in Graduate Archival Education Programs: Issues for Educators and the Profession”, *American Archivist* 63, núm. 2 (2000): 380-391.



1. **Un repositorio no es un *software*.** Un *software* no puede preservar nada, ni puede ser un repositorio en sí mismo. Un repositorio es la suma de recursos financieros, equipo de cómputo (*hardware*), tiempo del personal y la implementación continua de políticas y planificación para garantizar el acceso al contenido a largo plazo. Cualquier sistema de *software* que usted utilice para preservar y proporcionar acceso al contenido digital es necesariamente temporal. Es importante que pueda sacar su contenido de ahí, ya que lo más probable es que no dure para siempre. Del mismo modo, no existe ningún *software* que “haga” la preservación digital.

2. **Las instituciones hacen posible la preservación.** Cada uno de nosotros morirá. Sin cuidado y mantenimiento, las cosas que nos importaron persistirán durante un periodo de tiempo relacionado con la durabilidad de sus medios. Dicho esto, las principales facilitadoras de la preservación a largo plazo son nuestras instituciones (bibliotecas, archivos, museos, familias, organizaciones religiosas, gobiernos). Como tal, la posibilidad de la preservación se habilita a través del diseño y la función de esas instituciones. Sus organigramas, prácticas de contratación, financiamiento, credibilidad, etc., son todas partes clave de la maquinaria cultural que hace posible la preservación.

3. **Las herramientas pueden estorbar tanto como ayudar.** Existe la misma probabilidad de que las herramientas especializadas y el *software* para la preservación digital entorpezcan la resolución de los problemas o que ayuden a resolverlos. En muchos casos, es más fácil comenzar con poco e implementar herramientas y prácticas simples y discretas para realizar un seguimiento de la información digital utilizando sólo el sistema de archivos en el que usted trabaja. Es mejor comenzar de manera sencilla y luego introducir herramientas que le ayuden a mejorar el proceso a simplemente comprar algún sistema complejo sin haber organizado primero los procesos en casa.

4. **Nada ha sido preservado, sólo hay cosas que se están preservando.** La preservación es el resultado del esfuerzo continuo de las personas y la disponibilidad de recursos. El trabajo nunca está terminado. Esto es cierto para todas las formas de conservación; es sólo que los plazos para las acciones de preservación digital son mucho más cortos de lo que tienden a ser con la conservación de otras cosas, como libros o pinturas al óleo. Hay que evitar hablar sobre lo que “ha sido” preservado; sólo existe lo que “se está” preservando. Esto tiene ramificaciones significativas sobre cómo pensamos el trabajo del personal y los recursos para la preservación. Para saber



qué tan en serio se toma la preservación digital una organización, usted no debe comenzar observando su código y su arquitectura de almacenamiento o hablando con sus desarrolladores. Empiece charlando con sus encargados de finanzas y vea qué lugar ocupa la preservación digital en su presupuesto. Si una organización le da importancia a la preservación digital, se debe notar en la forma en que gasta su dinero. La preservación es un trabajo continuo. No es algo que pueda considerarse como un costo único.

5. La acumulación no es preservación. Es muy fácil comenzar a recopilar muchos objetos digitales y hacer copias de ellos. Eso no es preservación. Para preservar algo de verdad, usted debe asegurarse de que sea localizable y accesible, lo que requerirá que tenga un enfoque claro y coherente para el desarrollo de la colección, la organización, la descripción, el método y la estrategia para proporcionar el acceso a éste.

6. La copia de seguridad de datos no es preservación digital. Si usted comienza a hablar sobre la preservación digital y alguien le dice: “Oh, no hay que preocuparse por eso, se respalda todo por las noches”, debe estar preparado para explicar cómo y por qué el respaldo no cuenta como preservación digital. Este libro puede ayudarle a desarrollar esa explicación. Muchos de los aspectos que intervienen en la copia de seguridad de los datos para el uso normal son similares a los aspectos del trabajo de la preservación digital, pero las preocupaciones para poder restaurar los datos a mediano plazo son significativamente diferentes a las involucradas en garantizar el acceso al contenido en el futuro a largo plazo.

7. Los límites de los objetos digitales son borrosos. Los objetos individuales hacen referencia, incorporan y usan aspectos de otros objetos como parte de su función cotidiana. Usted puede creer que tiene una copia de una pieza de *software* al mantener una copia de su instalador, pero ese instalador bien puede llamar a un servicio web para comenzar a descargar archivos, en cuyo caso usted no puede instalar y ejecutar ese programa a menos que se tengan los archivos de los que depende. Es posible que necesite un conjunto de fuentes, un códec de video en particular o cualquier otro número de cosas para poder usar algo en el futuro, y es difícil distinguir con claridad lo que se encuentra dentro del objeto y lo que es externo a él.

8. La colección digital de una persona es el objeto digital de otra que es el conjunto de datos de otra. En algunos casos, el contenido de un disco duro se puede administrar como un solo elemento, en otros, se trata de una colección de elementos. En el mundo analógico, los límites de los objetos son un poco más claros o al menos se dan por hecho. Lo borroso de



los límites de los objetos digitales se refiere a que los conceptos “elemento” y “colección” son menos claros que su versión analógica. Por ejemplo, un sitio web puede ser un elemento en un archivo web, pero también puede ser una publicación periódica que cambia con el tiempo, así como una colección de archivos.

9. La preservación digital se trata de hacer el mejor uso de los recursos para mitigar las amenazas y los riesgos más apremiantes de la preservación. Nunca se termina con la preservación digital. No es algo que se pueda lograr o terminar. La preservación digital es un proceso continuo en el que se deben comprender los riesgos que pueden llevar a la pérdida del contenido o de la capacidad para reproducir e interactuar con él, y se deben utilizar los recursos que se tengan para mitigar esos riesgos.

10. La respuesta a casi cualquier pregunta sobre la preservación digital es “depende”. Los detalles importan en casi todos los casos. Decidir qué es lo que importa de un objeto o un conjunto de objetos depende en gran medida de cuál pueda ser su uso futuro. De la misma manera, la estrategia de preservación para una colección masiva y en constante crecimiento de videos de alta resolución será fundamentalmente diferente a la que tomaría una organización para garantizar el acceso a largo plazo a una colección de textos digitalizados.

11. Ya pasó la hora de comenzar a tomar medidas. Usted puede leer y reflexionar sobre modelos de datos complicados, esquemas para el seguimiento y registro de acciones para la preservación y una variedad de temas complejos e interesantes durante años, pero eso no ayudará a “mover las cajas del piso”, metáfora de la que hablaré más adelante. Hay cosas prácticas y funcionales que todos pueden y deben hacer ahora mismo para mitigar muchos de los riesgos de pérdida más apremiantes. He resaltado los puntos relacionados con la frase “mover las cajas del piso” en la segunda mitad del libro. Así que es necesario asegurarse de dar prioridad a estas cosas en lugar de profundizar en otras de las áreas de trabajo más abiertas de la preservación y la investigación digitales.

12. Las definiciones demasiado técnicas de la preservación digital son cómplices de silenciar el pasado. Gran parte del lenguaje y de las especificaciones de la preservación digital se han convertido en complejos conjuntos de requisitos que ofuscan muchas de las cosas prácticas que cualquier persona u organización puede hacer para aumentar la probabilidad del acceso al contenido en el futuro. Así, el poner a la preservación digital en



un marco muy técnico ha provocado que muchas instituciones más pequeñas y con menos recursos consideren que simplemente no pueden hacer preservación digital o que necesitan contratar consultores para que les expliquen los complejos estándares de metadatos de la preservación, cuando, en realidad, lo primero que deben hacer es una copia de sus archivos.

13. Las posibilidades² de los medios digitales generan la necesidad de que la preservación digital esté involucrada en el desarrollo de las colecciones digitales. Los medios digitales brindan oportunidades significativas y novedosas para involucrar a las comunidades en el desarrollo de las colecciones digitales. Cuando los encargados de la preservación digital dan por sentado que su trabajo sólo es preservar lo que se les da, no contribuyen a que dentro de una organización se pueda replantear lo que es posible recopilar. Las políticas de la preservación digital deben estar cada vez más conectadas e involucradas de manera directa con la política del desarrollo de colecciones. Es decir, las posibilidades de lo que se puede preservar con facilidad deben señalar las decisiones sobre lo que una organización quiere coleccionar y preservar.

14. La aceptación y acogida de la astilla del archivista. Nunca se ha guardado todo. Nunca se han guardado la mayoría de las cosas. Cuando se parte del entendimiento de que la mayoría de las cosas son temporales y es probable que se pierdan en la historia, se puede desviar y enfocar la energía para asegurar y procurar los recursos necesarios para proteger las cosas que más importan. La ideología de “lo digital” hace parecer que podríamos o deberíamos intentar guardar todo. Sin embargo, esto viene de la idea equivocada de que la preservación digital es sobre todo un reto técnico y no uno social y ético. Es necesario reconocer que existen diferentes niveles de esfuerzo que deberían encaminarse a proteger diferentes tipos de material en el futuro.

15. La escala y las estructuras inherentes de la información digital sugieren trabajar más con una pala que con una pinza. Si bien es necesario aceptar el hecho de que no se puede recopilar y preservar todo, también se debe reconocer que, en muchos casos, el tiempo y los recursos necesarios para tomar decisiones sobre cosas individuales podrían utilizarse mejor en otros lugares. Con frecuencia es mejor enfocar la toma de

2 N. del T.: en inglés se refiere al término *affordances*, es decir, las posibilidades que materialmente ofrece un objeto para reconocer de manera intuitiva para qué sirve o cómo usarlo; a lo largo del libro se utilizará la palabra “posibilidades”, en casi todos los casos, para denotar esta acepción.



decisiones de preservación digital a escala. Esto es particularmente cierto en los casos en que se trata de contenido que no es grande. De manera similar, muchas veces tiene sentido normalizar el contenido o procesar cualquier número de tipos de archivos derivados y conservar los originales. En todos estos casos, la computabilidad de la información digital y las realidades de los archivos digitales que contienen cantidades significativas de metadatos contextuales permiten ejecutar estas acciones por lotes y no de una en una.

16. El oficio de la preservación digital requiere pensar como un futurista. No se conocen las herramientas y los sistemas que las personas tendrán y usarán en el futuro para acceder al contenido digital. Entonces, si se quiere garantizar el acceso a la información digital a largo plazo, se debe, al menos en cierto nivel, pensar y estar al tanto de las tendencias en el desarrollo de las tecnologías digitales. Ésta es una consideración clave para la mitigación de riesgos. Los riesgos y las amenazas de la preservación se basan en los componentes tecnológicos actuales y del mañana, por lo que debemos mirar hacia el futuro de una manera en la que no se necesitaba hacer con los medios y formatos anteriores.

La preservación corresponde a las personas: notas sobre voz y estilo

Con esos 16 puntos fuera del camino, puedo aclarar algunas cuestiones sobre estilo y lenguaje para cerrar esta introducción. Dado que nada se preserva, sino que sólo existen cosas que algunas personas en alguna institución están preservando, un concepto básico de mi enfoque y mi forma de escribir es poner énfasis en los sujetos que están haciendo el trabajo. Gran parte del libro consiste en describir objetos individuales o colecciones que la gente que está en un contexto institucional se ha comprometido a preservar. Creo que es inapropiado descontextualizar la labor de la preservación digital. Como resultado de esto, a lo largo del texto he intentado mantener a esas personas y su trabajo como elementos para impulsar la narrativa. Hay que reconocer que gran parte del trabajo de la preservación digital, hasta la fecha, se ha situado como un campo altamente técnico al que se ha dedicado la élite de la élite de las instituciones culturales y que esto ha derivado en que muchas de esas historias traten de una clase privilegiada compuesta, en su mayoría, por hombres blancos que se dedican a preservar el trabajo y el legado de otros hombres blancos y privilegiados.



Dada la apremiante necesidad de que nuestra historia y nuestras colecciones representen de manera genuina la diversidad de nuestro mundo, he tratado de forma consciente de dirigir la atención hacia el trabajo que contrarresta este patrón. Espero que, al hacerlo, el campo pueda avanzar hacia enfoques más equitativos para recopilar, preservar y proporcionar acceso al pasado. Dicho esto, las historias de objetos y organizaciones no son tan diversas como me gustaría. Como mínimo, al reconocerlo por adelantado, quiero subrayar que los problemas que existen para hacer que nuestras instituciones reflejen la diversidad de las comunidades a las que sirven son muy reales y que, a pesar de los mejores esfuerzos, continuarán manifestándose, hasta cierto punto, en los ejemplos que describo.

Este énfasis en las personas, en las relaciones entre las personas, documentadas en las colecciones, y quienes hacen el trabajo para preservar y proporcionar acceso a la información digital, también se refleja en el lenguaje que uso para involucrar al lector. He intentado moldear este texto en los seminarios de posgrado de preservación digital que enseño. Entonces, cuando uso “usted” para referirme a usted, el lector, es importante aclarar qué suposiciones estoy haciendo sobre su persona. Asumo, por ejemplo, que quiere aprender sobre la preservación digital, lo que espero sea correcto dado que está leyendo este libro. En ese sentido, también supongo que tiene interés en participar en lo que llamo “el oficio de la preservación digital”. No estoy asumiendo que usted sea bibliotecario, archivero o curador, o que sea un estudiante que aspira a convertirse en uno. Ante todo, me imagino que el libro será útil para las personas que trabajan o quieren trabajar en instituciones culturales. Sin embargo, he intentado hacer que también le sirva a un público más amplio.

Espero que este libro sea de utilidad para los activistas que desean iniciar prácticas que garanticen el acceso a largo plazo a sus registros, para los académicos que se toman en serio el tratar de asegurar que su trabajo tenga una vida más allá de ellos, para aquellos que trabajan en grandes o pequeñas organizaciones sin fines de lucro o gubernamentales, y para quien tenga un interés particular en la historia de su familia o comunidad. Entonces, aunque espero que este libro le sirva a quienes desempeñan roles formales como custodios de la cultura y la historia, también aspiro a que este tema bastante complejo sea accesible para la mayoría de las personas. Hasta cierto punto, mi esperanza es que este replanteamiento de la preservación digital a un público más amplio pueda ofrecer algunas oportunidades potenciales para cambiar algunos de los problemas que he señalado antes con respecto al papel histórico de las clases privilegiadas como guardianas de la preservación.



Junto con esto, a lo largo del libro también doy por hecho un nivel mínimo de *rappport* entre usted y yo, cuando hablo de nosotros como “nosotros”, ya que al estar leyendo estas páginas, usted ha expresado al menos un poco de preocupación y deseo de garantizar el acceso a la información a largo plazo. Como tal, usted se ha integrado a una larga tradición mundial de conservadores de la memoria que forma parte de todas las culturas y sociedades humanas. Por lo tanto, “nosotros” pretende ser un término inclusivo que se refiere a cualquier persona que, de acuerdo con los valores de los bibliotecarios y archivistas, se comprometa a trabajar en un acceso ético, socialmente responsable y duradero a registros legibles de comunidades, culturas y sociedades.

Capítulo 1

Las tradiciones divergentes de la preservación

“La preservación digital no existe, sólo existe la conservación de los medios digitales”. Me han contado que ésta fue una de las frases con las que inició la reunión inaugural de un grupo de trabajo de preservación digital en la Library of Congress. El personal del programa de preservación digital de esa institución quedó estupefacto. Con tal comentario, parecía que el colega había descartado décadas de su trabajo y el de otros alrededor del mundo que gira en torno a la preservación digital. El esfuerzo de todo un año para definir la estrategia de preservación digital para la biblioteca más grande del mundo comenzó rechazando la idea misma de la preservación digital. A través de un diálogo profundo y sostenido, los miembros de este grupo de trabajo llegaron a entenderse y fueron capaces de desarrollar un marco institucional para la preservación digital. Esto, sin embargo, suscitó algunas preguntas importantes. ¿Cómo es posible que, en un área con décadas de trabajo y estudios, los colegas no pudieran reconocer de manera colectiva algunos de los conceptos más básicos del campo? Este capítulo ofrece una respuesta. Con demasiada frecuencia, un conjunto de supuestos centrales sobre lo que es con exactitud la preservación permanece sin expresarse.

He escuchado muchas historias como ésta, en las que el diálogo interdisciplinario sobre la preservación digital se rompe cuando un individuo protesta “pero *eso* no es preservación”. Las discusiones sobre este tema a menudo parten de la suposición tácita de que hemos resuelto lo que implica la preservación en el viejo mundo analógico. No lo hemos hecho. A menudo se da por sentado que antes de “lo digital” existía una noción particular, agradable y ordenada de lo que era la preservación, pero no es verdad. Esta idea equivocada es parte de una tendencia problemática general del furor digital que separa el universo en



cosas que vinieron antes y después de la computación. Antes de precisar qué es la preservación digital, resulta fundamental establecer su contexto entre la gama de linajes de preservación divergentes de siempre.

Preservación significa diferentes cosas en diferentes contextos. Las discusiones sobre preservación digital son particularmente polémicas ya que, en gran medida, estos diferentes contextos se han compactado. Los formatos en medios digitales han proliferado en casi todas las áreas en las que uno podría pensar en la preservación. Cada uno de esos contextos tiene una historia. Esas historias están ligadas a la naturaleza cambiante de los medios y los objetos para los que se desarrolló una concepción de preservación y conservación. Comprender un poco sobre las tradiciones divergentes de la preservación ayuda a establecer las nociones en competencia que están en juego para definir qué cuenta como preservación en qué contextos.

Identificar qué cuenta como preservación en determinada tradición o linaje es complicado. Como éste es un libro en cierta forma breve, intentaré simplificarlo. Con ese fin, he agrupado las tradiciones de preservación divergentes e históricamente contingentes en tres marcos distintos. Las describo como *artefactual*, *informativa* y *folclórica*. Éstas son correctas sólo en la medida en que son útiles. Lo que presento es menos una historia de la preservación que una movilización de esa historia para resolver cómo los conjuntos de ideas divergentes coexisten bajo la idea de preservación.

Para asegurarme de que este enfoque sea tan amplio como se necesite, exploro estas nociones de preservación en áreas tan diversas como el desarrollo de la preservación del entorno construido, la conservación del arte, la tradición manuscrita, los enfoques para la gestión de jardines de mariposas, los archivos sonoros y el desarrollo de la historia oral, así como desarrollo de la fotografía, la microfilmación y el reformateo para la preservación. Cada episodio y tradición ofrece un modelo mental para ser considerado y utilizado en diferentes contextos en la preservación digital. Cada caso proporciona un medio para probar y verificar nuestras propias suposiciones sobre qué es con exactitud la preservación. La comprensión de estos contextos y significados establece un vocabulario para articular qué aspectos de un objeto deben permanecer en el futuro para un propósito de uso determinado.

Las historias de la preservación están limitadas por una negociación entre lo que la gente quiere hacer y lo que el mundo material les permite hacer. En este contexto, la historia de los medios y formatos también debe entenderse como una contribución a nuestra conceptualización de la preservación. La materialidad y las posibilidades de nuestras formas, medios y plataformas



establecen los límites de lo que la preservación puede ser.³ Considero que las tradiciones de preservación se entienden mejor si las consideramos como la mitad del diálogo entre la naturaleza material y formal de las cosas que queremos preservar. Tomaré como base el trabajo que se ha hecho en la arqueología de medios, un campo de investigación en nuevos medios que busca comprender cómo su desarrollo ha moldeado el pensamiento y el conocimiento. Explicaré que los medios digitales deben entenderse como parte de un proceso continuo de remediación integrada en el desarrollo de una gama de nuevos medios y formatos que ofrecen un potencial de comunicación y preservación diferente. En este caso, debemos pensar ampliamente acerca de todo lo que da forma a las cosas que queremos preservar. Incluyendo, pero no limitando, las posibilidades de la palabra hablada, la memoria humana, los idiomas escritos, las tabletas de arcilla, las inscripciones en arenisca, el papiro, la vitela, las xilografías, los tipos móviles, las máquinas de escribir, los daguerrotipos, los dictáfonos, las películas, los gramófonos, los microfilms, los PDF, los CD-ROM y los archivos MP3. Dada la cacofonía entre medios, formatos y soportes, queda claro que la preservación no puede ser una sola área conceptual de trabajo. En cambio, la preservación se entiende mejor como un tipo de acto de malabarismo en el que los profesionales de la preservación, bibliotecarios, archivistas, curadores, conservadores, folcloristas, historiadores, arqueólogos, académicos, administradores de registros, antropólogos, numismáticos y filatelistas (entre otros), trabajan para llevar objetos folclóricos, informativos y de artefactuales al futuro. Pero me estoy adelantando; antes de entrar en temas de medios y soportes, es esencial articular estas tres tradiciones divergentes de la preservación.

¿La casa histórica, las palabras específicas o el meollo de la historia? Tres marcos para la preservación

Ésta *es* la finca Mount Vernon de George Washington. Éste *es* el *Frankenstein* de Mary Shelley. Ésta *es* la historia de Rama y Sita. Cada enunciado usa el mismo verbo, *es*, pero significa algo fundamentalmente distinto en cada contexto. Cada una de las tres declaraciones anteriores determina algo sobre la identidad de algún objeto que podríamos querer preservar. Cada declaración hace una afirmación diferente sobre lo que constituye la autenticidad o la integridad del objeto.

³ Para una primera aproximación a la relación entre las posibilidades de los objetos y la agencia de su interpretación y uso, ver Umberto Eco, *Kant and the Platypus. Essays on Language and Cognition*, trad. de Alastair McEwen (Nueva York: Harcourt Brace, 2000), 54.



En el primer caso, la identidad es *artefactual*, pues se basa en la noción de contigüidad histórica hacia un objeto físico tangible. Tiene que ser esa finca en específico para contar como la finca de George Washington. En el segundo caso, la identidad es *informativa*, basada en la idea de que cualquier copia con información idéntica es la misma. Cualquier ejemplar de *Frankenstein* servirá siempre que tenga las mismas palabras. El tercero es un caso de *folclor*, mientras los elementos clave de esa historia estén presentes, en realidad no importa si cada palabra o frase es la misma. Alguien puede contarla de manera diferente, pero, si las partes clave están allí, sigue siendo la misma. Cada uno de estos ejemplos proporciona un marco para diversas formas de preservación ancladas en las posibilidades de los objetos materiales.

La contigüidad histórica y el objeto artefactual

En el caso de la finca Mount Vernon, hablamos de un lugar, un sitio histórico, una plantación que fue propiedad del primer presidente de los Estados Unidos de América, a la orilla del río Potomac, que es administrada y mantenida por la Mount Vernon Ladies' Association. Para que la frase “Ésta es la finca Mount Vernon de George Washington” sea verdad, realmente deberíamos estar señalando la estructura. Si, en lugar de eso, sostengo una tarjeta postal de Mount Vernon haciendo la misma declaración, usted podría pensar algo como “Oh, quiere decir que es una foto de Mount Vernon. Sé que el edificio real no es la imagen”. Sabemos, por nuestra forma predeterminada de pensar, que una imagen de Mount Vernon no es Mount Vernon y que a lo que nos queremos referir es al lugar físico y a su estructura.

Entonces, ¿qué es lo que hace que Mount Vernon *sea* Mount Vernon? Como cualquier objeto físico, está cambiando en todo momento. La luz del sol interactúa con la pintura y la decolora. A medida que cambian las estaciones y las temperaturas varían, los materiales que componen la casa (madera, piedra, vidrio) se expanden y contraen a diferentes velocidades, con lo que ocasionan su desgaste. En el caso de este sitio histórico en particular (como muchos sitios históricos similares), se ha trabajado para restaurarlo con la intención de que luzca como lo hizo en un momento determinado. Incluso si no quisiéramos cambiar nada, la casa se deterioraría lentamente. La transformación constante hace que la esencia de Mount Vernon sea su contigüidad histórica. Aun si toda la madera, piedra y metal que forman la casa fuera reemplazada lentamente, seguiría siendo la casa en la que vivió George Washington. Es un artefacto y es un lugar.



Aquí no sólo importa que Mount Vernon sea un lugar específico. El sitio opera bajo la misma lógica con la que trabaja nuestra concepción de la identidad de una pintura como la *Mona Lisa*. Si le muestro una foto de la pintura y le digo “Ésta es la *Mona Lisa*”, lo más probable es que usted piense “Oh, quiere decir que ésta es una foto de la *Mona Lisa*, pero la *Mona Lisa* real está en París”. Debido a su enfoque en artefactos únicos, me he referido a esta concepción de la preservación como *artefactual*. En automático podríamos tener la misma noción *artefactual* de identidad para la Estatua de la Libertad, la Gran Esfinge de Giza o Angkor Wat. En cada caso, hay un artefacto singular o una constelación de artefactos que están en constante cambio y que contienen una cantidad infinita de información potencial para estudiar o explorar.

La ciencia de la conservación, el cuidado de las colecciones y la preservación histórica proporcionan métodos de preservación para los artefactos. La definición de preservación que en general se acepta en este dominio sostiene que “es una acción tomada para retrasar o prevenir el deterioro o daño al patrimonio cultural mediante el control de su entorno y/o el tratamiento de su estructura, para mantenerlos lo más cerca posible a un estado inmutable”.⁴ Este enfoque *artefactual* se centra únicamente en el cuidado de la manifestación física de un objeto. De acuerdo con este enfoque, los científicos de la conservación usan la ciencia de los materiales y la química para comprender cómo y por qué algún objeto se degrada e intentan intervenir para limitar el grado de deterioro. En este contexto, también pueden realizar tratamientos de conservación específicos y reparaciones para estabilizar un objeto y limitar su cambio físico.

En la conservación de obras de arte y la preservación de sitios históricos existe todo un vocabulario especializado bajo el concepto general de preservación. La restauración, por ejemplo, usa evidencias para restaurar un sitio o un objeto con el propósito de acercarlo a lo que alguna vez fue. Esto sin duda implica alterar o cambiar al objeto, pero, dado que siempre está cambiando, uno puede justificar cómo y por qué una acción particular de restauración justifica las intenciones por las que el sitio está siendo preservado. Junto con la restauración, la preservación histórica también ofrece la noción de “reconstrucción”. Ésta podría implicar la reconstrucción de un sitio desde cero con base en su supuesta ubicación histórica. En este caso, el sitio conserva su contigüidad histórica.

4 “The Conservator-Restorer: Definition of the Profession”, International Council of Museums, Committee for Conservation, acceso el 28 de abril de 2021, <https://www.icom-cc.org/en/definition-of-the-profession-1984>.



Las “mismas letras” y el objeto informativo

En el caso del *Frankenstein* de Mary Shelley, el verbo *es* quiere decir algo fundamentalmente distinto al *es* de la finca de Mount Vernon. No importa si encuentro la copia del libro que leí y subrayé en la secundaria o la que perteneció a mi madre, o la que era de Carl Sagan, todas satisfacen el requisito de ser el mismo libro. En el caso de los objetos de información, lo importante es que la información codificada sea idéntica. Mientras su copia y mi copia tengan las mismas palabras escritas de la misma manera que nuestro marco de referencia predefinido, todos podemos decir que tenemos el *Frankenstein* de Mary Shelley.⁵

Vale la pena señalar algunas de las diferencias fundamentales entre los objetos *informativos* y los *artefactuales*. En el caso de *Frankenstein*, el medio físico podría ser papel o podría leerse en un iPad, o podría estar escrito en las nubes por un avión; en cualquier caso, el medio es irrelevante. Mientras cada una de sus letras aparezca, la obra es auténtica. En general, haber leído *Frankenstein* es haber visto sus palabras en el orden correcto.

Esta lógica no sólo funciona para los textos escritos. Se puede decir que cada uno de los grabados de Albrecht Dürer *es* el grabado.⁶ Cada copia de una fotografía del mismo negativo *es* la fotografía. Cada LP del *White Album* de The Beatles *es* el álbum. Estos ejemplos demuestran un calificador “idéntico” en términos informativos. La foto, el grabado, el álbum, etc., representan manifestaciones de medios específicos que poseen la cualidad de tener las “mismas letras”. Para cada uno de estos medios tenemos una sensación que lo califica como informativamente idéntico. Por ejemplo, es probable que la mayoría de nosotros no estuviéramos de acuerdo en que una copia fotostática de baja calidad de una fotografía cuente como la foto. Por supuesto, la copia contiene parte de la información presente en la foto, pero no la suficiente como para contarla como igual. O para el caso, si alguien imprime mal una fotografía de un negativo, la imagen resultante no se consideraría como la foto. Por lo tanto, una característica clave del objeto *informativo* es su cualidad de ser igual. Cualquier

5 Mi distinción entre *informativo* y *artefactual* refleja las diferencias entre *alográfico* y *autográfico* que hace Matthew Kirschenbaum en *Mechanisms. New Media and the Forensic Imagination* (Cambridge, Massachusetts: MIT Press, 2008); al presentar a los estudiantes las dos últimas opciones, he encontrado que generalmente se confunden, por lo que he cambiado a estos otros términos que creo que son un poco más directos.

6 Un pequeño matiz: la primera impresión de un grabado es a menudo más precisa que las posteriores; en cuanto a la fotografía, de hecho, hay actos distintivos e interpretativos que uno puede hacer para crear una imagen diferente desde un mismo negativo.



aseveración acerca de que algo sea informativamente idéntico se deriva de una descripción tácita o explícita de qué es lo cuenta como información.

El concepto de “mismas letras” conlleva la idea de que existen ciertas condiciones de procedimiento para poder crear lo que constituye una instancia auténtica del objeto. Como otro ejemplo, considere las instalaciones lumínicas de Dan Flavin, en la década de 1960,⁷ que implicaban colocar una serie de luces en un espacio. Su obra en realidad era la luz, no las bombillas físicas que la creaban. Entonces, ¿qué constituye la obra en este caso? Cuando Flavin puso en venta sus instalaciones, no vendió conjuntos de bombillas, sino certificados de autenticidad que incluían diagramas sobre cómo armar las obras a partir de piezas estándar y la autorización para hacerlo en su nombre. Así, estableció los criterios para una instancia exitosa del trabajo, que contaban como las “mismas palabras” para estas instalaciones. Sin embargo, no especificó el tipo exacto de bombillas que utilizó (que ya no se producen), porque desde su perspectiva, como creador de la obra, las bombillas no formaban parte de la definición de lo que se requería para producirla auténticamente.

Los bibliotecarios han desarrollado un modelo conceptual sofisticado para describir estos diferentes niveles de identidad para los libros en un sistema llamado Functional Requirements for Bibliographic Records (FRBR). Vale la pena tomarse un momento para trazar este modelo sobre la forma en que he estado describiendo estas diferencias. En el sistema FRBR existe un libro tangible (el ítem), todas las copias de ese libro que son idénticas en términos de su función (la manifestación), el texto del libro en un idioma específico (la expresión) y la esencia de la creación de un autor (la obra).⁸ Gran parte del FRBR de manera particular está relacionado con la producción y difusión de libros, por lo que no se traslada de la misma forma a otros dominios, pero sí ilustra cuántos matices diferentes existen para diseccionar lo idéntico. Es significativo que sólo los dos primeros tipos de objeto (el ítem y la manifestación) son en realidad cosas físicas; existen como artefactos. Cuando la expresión funciona de manera *informativa*, se identifica no por su portador físico, sino por la información codificada en el portador. Dependiendo de con cuánta precisión se conceptualice la obra, bien podría ser *informativa* o *folclórica*.

7 Ejemplo tomado de Richard Rinehart y Jon Ippolito, eds., *Re-collection. Art, New Media, and Social Memory* (Cambridge, Massachusetts: MIT Press, 2014), 32-33; más allá de este punto específico, todo su libro está lleno de valiosas historias sobre cómo las obras de arte de los nuevos medios han desafiado los supuestos de la conservación del arte durante un siglo.

8 Barbara Tillett, “A Conceptual Model for the Bibliographic Universe”, *Technicalities* 25, núm. 5 (2003), <http://www.loc.gov/cds/downloads/FRBR.PDF>.



A lo largo de este capítulo me he referido a cada una de estas categorías como marcos de referencia predeterminados. Habiendo explicado un poco sobre dos de ellos, puedo desarrollar más el concepto. No es que estas concepciones de preservación sean inherentes a algún objeto determinado, pues en cualquier momento podemos abordarlo como artefactual o informativo. En el caso de los objetos informativos, sólo pueden existir si se codifican en algún medio. Entonces, todos los objetos informativos deben tener algo así como un anfitrión artefactual en el que habitan. De manera similar, como sugerí anteriormente, cada objeto artefactual alberga en sí mismo una cantidad prácticamente infinita de información. Algunos ejemplos pueden ayudar a concretar estas ideas.

Uno de ellos es el de los primeros libros impresos en inglés moderno temprano, como las obras de Shakespeare. Éstos son útiles para estudiar esta intersección, ya que tienen propiedades informativas claras que establecen el criterio de “las mismas letras”, pero también se han estudiado cada vez más como cultura material.⁹ Usted puede encontrar copias en blanco y negro del segundo cuarto de *Hamlet* en recursos en línea como Early English Books Online. Esas copias, sin embargo, fueron escaneadas de un microfilm de baja calidad tomado del original. Cuando usted observa el microfilm digitalizado, puede ver la información, las letras que componen las palabras, su diseño y composición en la página. No obstante, la copia digital es un poco engañosa, ya que el ítem original también incluía información esencial codificada en un color rojo más claro que no es visible en el microfilm blanco y negro. En gran medida, esto podría resolverse sólo con digitalizar mejor las páginas del libro. Por fortuna, la Folger Shakespeare Library sí proporciona mejores digitalizaciones a mayor resolución. Dicho esto, usted también puede imaginar cómo alguien que realiza la codificación textual podría usar el marcado para indicar qué texto es de qué color. En efecto, algunos trabajos recientes con manuscritos medievales han demostrado que hay formas de obtener mucha más información almacenada en estas obras.

No sólo se trata de la información codificada intencionalmente, a la que podríamos acceder y conservar para el futuro. Resulta sorprendente saber que también existe una cantidad considerable de información valiosa que se encuentra en “la suciedad” de las páginas que dejaron ahí las personas

⁹ Los ejemplos de Shakespeare discutidos aquí provienen de Sarah Werner, “Where Material Book Culture Meets Digital Humanities”, *Journal of Digital Humanities* 1, núm. 3 (2012), <http://journalofdigitalhumanities.org/1-3/where-material-book-culture-meets-digital-humanities-by-sarah-werner/>.



que lo han tenido en sus manos a través de los años. Utilizando densitómetros e imágenes digitales de copias de manuscritos de muy alta calidad, los académicos han podido estudiar la relativa suciedad de sus páginas individuales, lo que a su vez nos cuenta una historia sobre el uso de los objetos. De la misma manera, es posible imaginar que se realice el análisis de las propiedades químicas del pegamento en la encuadernación de un libro o se observe la celulosa en sus páginas para descubrir con qué árboles crearon el papel. Con toda seriedad, existe un interés continuo en estudiar el olor de diferentes libros modernos tempranos, ya que podría ayudar a demostrar aspectos de su circulación. Los olores poseen propiedades químicas que pueden registrarse. Todo lo mencionado tiene el propósito de enfatizar que, si bien tiene sentido decidir centrar las actividades de la preservación en las cualidades informativas de un objeto, ello conlleva siempre afirmaciones implícitas o explícitas respecto a qué aspectos del material subyacente y del objeto artefactual necesitan ser conservadas. Las copias individuales de las obras de Shakespeare tienen propiedades informativas, pero también existen como artefactos raros y únicos por derecho propio.

La misma dualidad de las propiedades informativas y artefactuales de los objetos está presente en las obras que usamos de manera predeterminada en un contexto artefactual. Todos los artefactos contienen información, y ésta siempre está implicada en las cualidades artefactuales del medio en el que está codificado. Uno de los propósitos principales por los que mantenemos, conservamos y preservamos artefactos es su potencial valor informativo. Si bien pensamos por defecto en el marco artefactual para cosas como pinturas o esculturas, la mayoría de las formas en que realmente trabajamos con esos objetos son representaciones informativas subrogadas de ellos. Las colecciones de diapositivas que han sido un pilar en la educación de la historia del arte son parte de cómo los estudiantes identifican las obras y las tendencias en esta disciplina. Si bien es justo decir que usted no ha visto en verdad la *Mona Lisa* hasta que la mire en la pared del museo de Louvre, cuando se trata de enseñarle a la gente sobre la historia del arte, no insistimos en que vayan y vean cada objeto individual. En cambio, tendemos a estar bien con el uso de fotografías que documentan las características informativas clave de esas obras para usarlas en la enseñanza (aquí la noción de documentación abre una idea diferente en la preservación a la que volveremos más adelante). Del mismo modo, cuando los académicos publican en una revista de historia del arte, a menudo incluyen imágenes de la obra que están discutiendo y analizando. No dicen: “toma mi artículo y ve a ver cada una de las obras de las que hablo”. Las imágenes hacen



que las cualidades informativas con las que quieren trabajar sean legibles y, por lo tanto, son suficientes para el propósito de su ensayo. Al reproducir estas imágenes lado a lado, el historiador puede dirigir la atención hacia ciertas características del objeto e invitar a la comparación e interpretación, lo que no es posible cuando se visitan las obras físicamente.

Entonces, no es que existan formas innatas de preservación asociadas a ciertos tipos de objetos. Lo que hay, en cambio, son marcos de referencia distintos para abordar cualquier objeto dado, algunos de los cuales tienden a ser los valores predeterminados para la preservación en un área determinada, con un medio o soporte específico. Hasta ahora he descrito dos de esos marcos (*artefactual* e *informativo*). Presentaré uno más y luego trabajaré con una serie de contextos para aplicarlos.

El “mismo significado” en los objetos folclóricos

La historia de Rama y Sita, como se cuenta en el épico poema sánscrito del Ramayana, nos fue explicada a mi esposa y a mí por un guía turístico balinés como “nuestro Romeo y Julieta”. En la historia, el príncipe Rama, ayudado por un ejército de monos, viaja para rescatar a su esposa Sita del malvado Ravana. Fechada entre 500 a. C. y 100 a. C., la historia existe con cientos de variantes y es fundamental para las múltiples tradiciones religiosas y culturales agrupadas por el hinduismo. La historia no sólo existe en estos textos, también en más o menos todos los medios culturales.

Si bien hay algunas partes fundamentales en la narrativa, que son cruciales para presentar la historia, la variabilidad de su expresión en diferentes culturas y medios es, en sí misma, importante para su esencia. En algunas versiones, la naturaleza humana de Rama es central, en otras, su divinidad. Incluso, en algunos casos, Sita es un personaje más importante. La variante principal de Java incluye una gama de otras deidades regionales. Estas versiones de la misma historia no satisfacen el criterio de las “mismas palabras”, para ser informativamente idénticas; no obstante, si usted le preguntara a alguien si todas cuentan como la historia de Rama y Sita, existe una buena posibilidad de que la respuesta sea sí. Se puede considerar a cierto nivel que esto satisface una concepción mucho más flexible o más difusa del requisito informativo de las “mismas palabras”. La historia necesita tener algunos elementos o características clave para contar como la misma, pero estos requisitos son mucho menos restrictivos de lo que hubiera pasado en el caso de un marco informativo. Un etnógrafo o un folclorista sin duda diría que estas variaciones en verdad son la historia. De hecho, es probable que sugirieran que la variabilidad de estas



narraciones es una de las cosas más importante sobre ellas. Los especialistas sostienen que, cuando el folclor se fija en un medio, pierde algo de la vitalidad cultural que existe en su manera de mutar y variar. En cierto sentido ya no es cultura viva, sino algo que ha sido inmovilizado. Comprender un poco sobre la manera en que los folcloristas entienden la cultura será útil para ver cómo esto representa sólo un tercio, nuevamente, de una tradición en principio distinta en la preservación.

Existe un malentendido generalizado de que el folclor se refiere a “viejas historias” o cuentos. En la práctica, se trata más de la lógica y el significado de lo cotidiano. El folclor trata sobre lo informal y lo de todos los días, sobre cómo se transmiten las tradiciones y las costumbres.¹⁰ No sólo las historias presentan estas cualidades, también las cosas que decimos (chistes, canciones, mitos y leyendas), las cosas que hacemos (costumbres, juegos, rituales, ritos de iniciación), las cosas que fabricamos (artesanías hechas a mano, colecciones personales) y las cosas que creemos (supersticiones, ideas sobre lo sobrenatural, lo vernáculo y la religión). Con este fin, los folcloristas hacen todo lo posible para buscar variaciones de la forma en que se cuenta una historia en la vida cotidiana y trabajan para documentarlas, con el fin de rastrear cómo emergen y cambian, adaptándose a las necesidades de un momento determinado. Fuera de estas prácticas formales de recopilar el folclor, las culturas y las sociedades humanas sirven como un medio para preservarlo en aquello que es útil para sus culturas. Los folcloristas pueden documentar y estudiar folclor y, de hecho, crean objetos informativos que se conservan de acuerdo con su marco. Sin embargo, las sociedades refinan y revisan de manera continua qué aspectos del conocimiento popular son útiles y llevan esa información a través de la vida cotidiana y las prácticas de las personas.

Esta forma *folclórica* de preservación es un aspecto fundamental de todas las sociedades humanas. Es la forma más antigua de preservación. Es más, se trata de uno de los factores esenciales que sostienen y hacen posible la civilización desde sus orígenes. Es la base de la cultura y el idioma.¹¹ Piense en todo

¹⁰ Mis descripciones respecto al folclor en gran parte están basadas en Lynne S. McNeill, *Folklore Rules. A Fun, Quick, and Useful Introduction to the Field of Academic Folklore Studies* (Logan: University Press of Colorado, 2013).

¹¹ Michael Tomasello describe los aspectos folclóricos de la memoria sociocultural como el “engranaje cultural”, un mecanismo social que permite que las ideas y los conceptos desarrollados por los individuos se difundan a través de una comprensión compartida de ver a los demás como personas iguales a nosotros mismos, Michael Tomasello, *The Cultural Origins of Human Cognition* (Boston, Massachusetts: Harvard University Press, 2009), 5.



lo que sabemos que no aprendimos leyendo o en algún tipo de medio. Cuando aprendemos el idioma a través de nuestros padres o nuestras familias, este proceso se pone en marcha. Así, el “soporte” del folclor es la memoria agregada, encarnada y representada por una sociedad determinada. A medida que las tradiciones y las prácticas se recrean y se repiten, difunden información hacia los demás. En este sentido, la preservación *folclórica* ha funcionado como una parte central de las culturas humanas, al menos desde los inicios del *Homo sapiens*, hace unos 200 mil años. Existe una buena posibilidad de que podría haber funcionado como parte de las sociedades neandertales o de humanos arcaicos, anteriores a ellos, desde hace dos millones de años. Vale la pena subrayar esta historia, en parte, porque el marco *folclórico* es probablemente el más crítico de los tres, uno que en esencia está codificado en nuestros cuerpos. Sin embargo, este modo de preservación suele aparecer con menos frecuencia o tiene menor probabilidad de ser parte de la mayoría de las definiciones formales de preservación y conservación. Es tan fundamental para lo que somos que a menudo estamos demasiado cerca como para hacerlo consciente.

Durante el siglo xx, los folcloristas recurrieron cada vez más a los nuevos medios para documentar las variaciones que existen en la naturaleza performativa del folclor. En este sentido, utilizan registros sonoros, fotografías y películas para capturar expresiones individuales de la experiencia y la perspectiva. En estos casos, el surgimiento de esos nuevos medios hizo posible un tipo completamente nuevo de preservación y, al mismo tiempo, representó desafíos importantes en cuanto a la preservación y conservación.

No es exagerado afirmar que estos, ahora viejos, medios (fotografía, película, grabaciones sonoras, etc.) dieron origen a nuevas ideas sobre lo que podría ser la preservación. En 1877, un artículo de la revista *Scientific American* afirmaba que gracias a la invención del fonógrafo “el discurso se ha vuelto, por así decirlo, inmortal”.¹² Es difícil imaginar lo poderoso que debe haberse sentido reproducir el sonido real de un evento pasado por primera vez. Es significativo que el impacto de la fonografía (cuyo significado literal es “escribir sonidos”, en griego) cambió las nociones de lo que era la memoria. De forma parecida a la publicidad de computadoras en las últimas décadas, que compara nuestros cerebros con discos duros, un artículo de 1880 sobre la memoria y el fonógrafo, define “al cerebro como un fonógrafo infinitamente perfeccionado”.¹³ El

12 “A Wonderful Invention—Speech Capable of Indefinite Repetition from Automatic Records”, *Scientific American* (17 de noviembre de 1877): 2.

13 Jean-Marie Guyau, “Memory and Phonograph” [1880], citado en Friedrich A. Kittler, *Gramophone, Film, Typewriter*, trad. de Michael Wutz y Geoffrey Winthrop-Young (Stanford, Ca-



desarrollo de este aparato, y de otros medios nuevos a lo largo del tiempo, demostró cómo las posibilidades de nuestros propios cuerpos forman parte de este diálogo. Replanteamos nuestras concepciones de la memoria junto con lo que nuestros medios nos permiten, construimos nuestros medios en el imaginario de nuestras mentes y, por el contrario, imaginamos nuestras mentes a través de nuestros medios, como fonógrafos, cámaras de video y discos duros.

El desarrollo de las grabaciones sonoras generó nuevas alternativas para preservar y grabar el mundo y el patrimonio cultural, y los profesionales de ese ámbito lo aprovecharon muy pronto. Desde la década de 1920, los cilindros de cera se usaban para recopilar y preservar canciones populares.¹⁴ En este contexto, el campo de la historia oral, la recopilación intencional y el registro de historias personales se desarrollaron junto con estas tecnologías. En la primera reunión del National Colloquium on Oral History en 1966, se sugirió que la “grabadora de cinta [era] lo suficientemente importante para la historia oral como para constituirse como una parte de su definición”.¹⁵ Los nuevos medios no sólo modifican nuestras nociones individuales de mente y memoria, sino que también dan forma y modelan el desarrollo de profesiones enteras, como la historia oral, que funcionan como órganos de la memoria cultural de las sociedades. El surgimiento y la difusión de estas tecnologías de grabación permitieron nuevos tipos de trabajo académico y de preservación. Sin embargo, no fue tan claro el impacto que tendría la tecnología en estas nuevas prácticas. Las prácticas del folclor han estado históricamente ancladas en la investigación textual. La introducción de tecnologías de grabación generó controversias sobre cuál debía considerarse como la fuente real o el objeto a preservar. Para muchos, la transcripción de la grabación era y sigue siendo el artefacto principal, mientras que otros han luchado para que el registro original en bruto sea reconocido como tal. Las nuevas tecnologías de medios no determinan en sí mismas el futuro de la preservación. Las posibilidades de esos medios entran en el diálogo continuo con la práctica profesional donde se negocia su posición.

lifornia: Stanford University Press, 1999), 33; estos ejemplos son fundamentales para el argumento que Kittler desarrolla en esta obra como parte de la arqueología de los medios y también desempeñan un papel central en la obra de Lisa Gitelman, *Always Already New. Media, History, and the Data of Culture* (Cambridge, Massachusetts: MIT Press, 2006).

14 Para una exploración de las intersecciones entre los nuevos medios y la preservación del folclor, ver Nicole Saylor, “Computing Culture in the AFC Archive”, *Folklife Today* (blog), 8 de enero de 2014, <https://blogs.loc.gov/folklife/2014/01/computing-culture-in-the-afc-archive/>.

15 Se puede encontrar una descripción de esta reunión en Rebecca Sharpless, “The History of Oral History”, en *History of Oral History: Foundations and Methodology*, ed. de Thomas L. Charlton, Lois E. Myers y Rebecca Sharpless (Lanham: AltaMira, 2007), 15.



A medida que diferentes viejos nuevos medios se han desarrollado (cilindros de cera, discos, cintas de casete, películas de ocho milímetros, cintas VHS), han sido considerados como medios para registrar y transportar la señal, la grabación o la imagen. Es decir que el marco de referencia predeterminado para estos materiales documentales es el informativo. Dado que muchos de estos medios audiovisuales no se caracterizan por ser longevos, los profesionales del campo de la preservación audiovisual han llegado a un consenso sobre la necesidad de digitalizar este contenido, en particular el que viene en cinta, como un medio de preservación.¹⁶ Este enfoque concuerda por completo con entender ese tipo de contenido en términos informativos.

Vale la pena señalar de manera breve que la documentación ha sido un modo primario de preservación para las artes escénicas (teatro, danza, música). Parte del mismo principio que los estudios del folclor. Una representación (actuación) es un único evento temporal vivido. No persiste en el tiempo y, aunque se puede tomar algo como una partitura musical o el guion de una obra de teatro y hacer copias perfectas de su información, no sustituirá la experiencia vivida de haber estado en el evento. Así, de manera similar al folclor y a la historia oral, el campo de la preservación de las artes escénicas ha adoptado y adaptado los nuevos medios, pues han surgido como una forma de grabar y proporcionar acceso a las representaciones de obras. El caso de la documentación de las artes escénicas aclara algunas de las tensiones clave en el corazón de la preservación.¹⁷ Si bien podemos preservar cosas, objetos tangibles, a menudo nos preocupamos mucho por las experiencias. La única forma de preservar esas experiencias es, de alguna manera, registrando aspectos de ellas en un medio. De forma similar, el enfoque para este tipo de documentación también está basado en un marco de referencia informativo. Los medios son elegidos por los documentalistas y, por lo tanto, carecen de mucha de la información artefactual asociada a los objetos en los que el artefacto original es, en sí mismo, parte de la elección de los creadores.

16 Sam Brylawski *et al.*, eds., “ARSC Guide to Audio Preservation”, *CLIR Publication* (Washington DC: Association for Recorded Sound Collections / Council on Library and Information Resources / The Library of Congress, 2015), <http://cmsimpact.org/wp-content/uploads/2016/08/ARSC-Audio-Preservation.pdf>.

17 A la fecha, un recurso fantástico para comprender la documentación de las artes escénicas sigue siendo la obra de Libby Smigel, Martha Goldstein y Elizabeth Aldrich, *Documenting Dance. A Practical Guide* (Washington DC: Dance Heritage Coalition, 2006), <https://web.archive.org/web/20180417151147/http://www.danceheritage.org/DocumentingDance.pdf>.



Epopeyas, archivos y jardines de mariposas: la aplicación de los marcos

He introducido tres marcos diferentes para pensar respecto a la preservación: lo artefactual, lo informativo y lo folclórico, y he sugerido la manera en que los distintos tipos de objetos se clasifican en estos marcos conceptuales predefinidos. En ese proceso, he intentado demostrar cómo se conectan y se cruzan los diferentes marcos y cómo pueden aplicarse a la preservación en un contexto dado. Para avanzar más en este punto, tomaré tres ejemplos adicionales: la *Iliada*, registros de archivo en general y diferentes tipos de colecciones de mariposas como contextos adicionales para mostrar la forma en que estos diferentes marcos de preservación se unen para permitir el acceso futuro a objetos, información e historia. Al hacerlo, tengo la intención de presentar una serie de cuestiones y conceptos que son útiles para pensar en lo que podemos y debemos significar cuando, finalmente, lleguemos a la preservación digital.

La *Iliada* de Homero

La *Iliada* es un ejemplo útil para explorar las interconexiones entre los marcos de preservación que he descrito. En particular, el hecho de que esta obra sea anterior al lenguaje escrito en Europa y continúe siendo una historia esencial, así como un registro histórico importante para gran parte de la civilización occidental, nos ofrece un contexto rico para explorar cómo algo puede perdurar a través del tiempo. Como tal, estas historias son una forma eficaz de ilustrar la idea de que los diferentes hilos de preservación se cruzan. Como con todos mis ejemplos, éste no pretende ser un estudio profundo del tema, sino una oportunidad para movilizar parte de la historia para pensar en la interacción de algunos de los conceptos que propongo. Entonces, pido disculpas de antemano a los expertos por cualquier detalle que no haya presentado.

La *Iliada* probablemente fue compuesta entre los siglos VII y VIII a. C. Cuenta las historias de eventos que se piensa que ocurrieron unos 400 años antes, a principios del siglo XII a. C. El primer manuscrito completo al que tenemos acceso fue copiado en el siglo X d. C. Como tal, esta obra persistió en la tradición oral y luego en una larga historia de tradición manuscrita. Muchas de las características clave del texto reproducen aquellas que se vinculaban a las posibilidades del medio de la memoria humana para volverlas a contar. En particular, utiliza patrones de repetición que fueron esenciales para que las narraciones épicas funcionaran en una sociedad preliteraria. En este sentido, usted puede pensar en las versiones escritas de las obras como la remediación de



su forma original. De la misma manera en que las primeras películas fueron grabaciones funcionales de representaciones escénicas, las obras escritas más tempranas, como ésta, acarrearán elementos de las posibilidades de los medios anteriores hacia las formas de los posteriores. Entonces tenemos esta transición de la tradición oral folclórica a la replicación informativa de los textos escritos.

Muchos lectores, incluido el autor, experimentaron la historia de la *Iliada* sólo mediante traducciones. Sin embargo, éstas, en términos generales, son interpretaciones basadas en más de una fuente textual. Existen más de dos mil copias manuscritas de las obras de Homero y un campo académico entero que explora las variaciones y cambios entre ellas para comprender mejor cómo se desarrolló y cambió esta historia a lo largo del tiempo.¹⁸ El problema es que no tenemos copias del texto de cuando es más probable que se haya escrito originalmente. Lo que tenemos es una serie de copias distintas fechadas en diferentes momentos. Mientras que la tradición manuscrita opera a través de una lógica de identidad informativa en la que cada copia es auténtica si tiene las mismas letras, en la práctica, el proceso de creación de copias de un texto introduce variaciones a lo largo del tiempo. Los humanos simplemente no somos copistas perfectos. Las desviaciones que surgen con el tiempo y las variaciones en los textos parecen un poco más folcloristas que informativas.

Entonces, ¿cómo se trabaja con estas copias? Es importante resaltar aquí que las características artefactuales de cada una de esas copias se vuelven cruciales, es decir, todo el contexto que gira en torno al hallazgo de una copia, la información sobre el material del que está hecha y su datación por carbono, se convierten en la evidencia esencial utilizada para reconstruir la historia de estos textos. Cabe destacar que las cualidades del artefacto, en particular la naturaleza duradera de los medios en los que se codificaron estos textos, son lo que nos ha permitido obtener acceso a su contenido informativo. Entonces, los tres marcos han sido fundamentales para lograr una buena copia del trabajo que podemos usar como fuente para todo tipo de cosas. Es preciso subrayar que la historia de cada uno de esos aspectos se convierte en una parte clave para pensar en las aportaciones que nos puede brindar el texto. En el caso de las obras de Homero, los historiadores han derramado considerable tinta a lo largo de los años para argumentar hasta qué punto pueden proporcionar evidencia de la historicidad de la ciudad y la Guerra de Troya.

18 Graeme D. Bird, *Multitextuality in the Homeric Iliad. The Witness of the Ptolemaic Papyri*, Hellenic Studies 43 (Washington DC; Cambridge, Massachusetts: Center for Hellenic Studies / Harvard University Press, 2010).



Como nota final sobre la *Iliada*, también vale la pena enfatizar que las historias son mucho más extensas que los textos, pues éstos existieron en un momento determinado. En otras palabras, el proceso folclórico mediante el que una narración se reinterpreta y explora en formas expresivas nuevas e innumerables continúa a buen ritmo. Estas historias han sido remediadas como obras de teatro, ballets, libros para niños, cómics, películas, etc. Todas estas encarnaciones son funcionales para conservar partes de la historia. Además, a medida que las historias se transforman y mutan en nuevas versiones, también descubrimos qué es lo que sigue siendo suficientemente vital para nuestra cultura para que la sociedad lo preserve en nuevas formas. Es posible que estas variaciones nos enseñen tanto sobre lo que significan las historias como la copia más prístina y auténtica que podamos reunir en la mejor edición crítica de los textos.

Los registros de archivo

Sería negligente escribir sobre las tradiciones de la preservación sin discutir sobre los archivos. La producción y gestión de registros es una de las tradiciones de preservación más antiguas y es un buen lugar para destacar algunas de las intersecciones que hay entre nuestros diferentes marcos de preservación.

Muchos de los objetos preservados más antiguos de la historia humana tienen forma de registro. La creación de estos últimos es, en sí misma, parte de la base del desarrollo de los sistemas de escritura. Las tabletas cuneiformes babilónicas eran utilizadas para registrar inventarios en graneros. De la misma manera, en la actualidad se conservan registros de observaciones astronómicas del antiguo Egipto y China que todavía se utilizan como parte de los datos con los que trabajan hoy los astrónomos. En estos dos ejemplos, podemos apreciar cómo diferentes marcos de preservación han llegado a desempeñar un papel importante. Las tabletas cuneiformes resultaron ser uno de los medios más duraderos para codificar información, dado que han sobrevivido a una cantidad considerable de negligencia. Por el contrario, los registros astronómicos de todo el mundo han persistido en gran medida gracias a que su información era tan útil y vital que de manera continua se copió y migró hacia nuevos medios.

La práctica de la archivística moderna es bastante pragmática en términos de la preservación de sus registros. Se enfoca tanto en mejorar las condiciones de almacenamiento para evitar la descomposición de los objetos materiales en los que los registros están codificados como en mantener técnicas de apoyo, como la microfilmación, que crea copias de documentos en microfilms que, en condiciones de almacenamiento adecuadas, podrían durar casi mil años. En esta duali-



dad de estrategias podemos ver cómo se toma en cuenta tanto el objeto material como el informativo. Sin embargo, la atención que se da al objeto material está más enfocada en tratar de asegurar la longevidad de la información almacenada en él de manera económica que en sus cualidades artefactuales.¹⁹

El enfoque en el objeto informativo también es evidente en la tradición de la preservación de registros conocida como “edición documental”. Aunque usted puede pensar que se refiere a las películas documentales, en realidad este campo se enfoca en la edición y publicación de las ediciones de los registros textuales. La Association for Documentary Editing señala que el proyecto Papers of Thomas Jefferson comenzó en Princeton en 1943 y provocó, en parte, que los United States National Archives crearan la National Historical Publications Commission.²⁰ Los proyectos de edición documental conservan los registros mediante la publicación, por ejemplo, transcribiendo cartas manuscritas de manera que puedan ser impresas y compartidas como texto. Vale la pena destacar que esta estrategia para publicar registros es mucho más antigua que la edición documental moderna. Los esfuerzos detrás del *Thesaurus Graecae Linguae* se remontan a la obra del erudito editor francés Henri Estienne en el siglo XVI.²¹ En muchos casos, las únicas copias existentes de varias fuentes primarias perviven gracias a su publicación en estos volúmenes.

En este contexto, podemos empezar a ver la publicación de registros a partir de la llegada del tipo móvil como una continuación de la muchísimo más antigua tradición manuscrita en la que los escribanos copiaban los textos de forma meticulosa, más o menos desde el advenimiento de la escritura misma. Esto subraya el proceso continuo mediante el cual los interesados en la preservación de la información han explorado y explotado las posibilidades de los nuevos medios. Cuando la tecnología del microfilm se desarrolló en la década de 1930, los bibliotecarios, los archivistas y los historiadores estaban encantados con la posibilidad de reproducir las fuentes.²²

19 Para una discusión más detallada de estos puntos, ver el segundo capítulo de la obra de David Bearman, *Archival Methods*, Archives and Museum Informatics Technical Report, vol. 3, núm. 1 (Pittsburgh, Pensilvania: Archives & Museum Informatics, 1989).

20 “About Documentary Editing”, Association for Documentary Editing, acceso el 28 de abril de 2021, <http://www.documentaryediting.org/wordpress/?pageid=482>.

21 La historia del *Thesaurus Graecae Linguae* original se describe de manera breve en el contexto de un trabajo reciente que amplía y desarrolla aún más esta visión, ver “Thesaurus Linguae Graecae—History”, Thesaurus Linguae Graecae Project, acceso el 28 de abril de 2021, <https://web.archive.org/web/20120129001615/http://www.tlg.uci.edu/about/history.php>.

22 El historiador público Ian R. Tyrrell usó la misma retórica comúnmente utilizada sobre la digitalización y la red para describir el microfilm en los años 30 del siglo XX, en sus palabras, el



Aunque el enfoque en el contenido informativo de los registros ha sido y sigue siendo primordial, siempre hay una cantidad casi ilimitada de información en cualquier objeto material que podría ser estudiada y explorada para proporcionar datos adicionales. El trabajo reciente en imágenes multiespectrales ha sido particularmente relevante en esta área. Mediante la creación de imágenes compuestas de registros, como el borrador de la Constitución de los Estados Unidos o del *Palimpsesto de Arquímedes*, estas técnicas han demostrado su viabilidad para descubrir información que parece oculta en el interior de las características materiales o artefactuales de un registro de archivo. En el primer caso, dichas técnicas fueron centrales para demostrar que sin duda alguna Thomas Jefferson tachó el término “sujetos” y lo reemplazó por el de “ciudadanos”.²³ En el segundo, una serie de obras de Arquímedes que se creían perdidas en el tiempo fueron recuperadas a partir de un libro de oraciones del siglo XIII. Las obras del filósofo habían sido borradas y el pergamino reutilizado, pero, a través de las imágenes multiespectrales, los trazos de los textos que son invisibles al ojo humano pudieron aparecer y hacerse legibles.²⁴ Estos puntos de las nuevas técnicas de la imagen demuestran los desafíos de articular los aspectos informativos de los registros. Los artefactos contienen información potencial casi ilimitada. Además, estos ejemplos muestran que el desarrollo de nuevos tipos de tecnologías no sólo continuará generando nuevos medios, sino que también contribuirá con nuevos métodos y técnicas para mostrar los aspectos informativos en los objetos artefactuales.

Formas de preservar las mariposas

Hasta ahora, todos mis ejemplos han sido objetos u obras generados por humanos. Es factible probar la eficacia de los marcos de la preservación al explorar qué tan bien funcionan en relación con la preservación en el mundo natural.

microfilm “democratizó el acceso a las fuentes primarias en la década de 1960 y, por lo tanto, le dio un valor agregado a la investigación original y a los enfoques monográficos”, Ian R. Tyrrell, *Historians in Public. The Practice of American History: 1890-1970* (Chicago: University of Chicago Press, 2005). La tabla de contenido de la obra está disponible en <http://www.loc.gov/catdir/toc/ecip058/2005003459.html>.

²³ “Hyperspectral Imaging by Library of Congress Reveals Change Made by Thomas Jefferson in Original Declaration of Independence Draft” [2 de julio de 2010], Library of Congress Office of Communications, acceso el 28 de abril de 2021, www.loc.gov/item/prn-10-161/analysis-reveals-changes-in-declaration-of-independence/2010-07-02/.

²⁴ Para más información sobre el proyecto Archimedes Palimpsest, ver “About the Archimedes Palimpsest”, Archimedes Palimpsest Project, acceso el 28 de abril de 2021, <http://archimedespalimpsest.org/about/>.



En este contexto, contrastar diferentes tipos de colecciones de mariposas ofrece una forma útil de ver cómo se desarrollan estos conceptos.

La sala principal del Milwaukee Public Museum ofrece un “jardín de mariposas” en un extremo y al otro una exposición de cajas de madera que contiene una colección de especímenes preservados.²⁵ Estos dos lados opuestos de la sala son útiles para contrastar los marcos de la preservación. También demuestran la manera en que éstos funcionan de forma similar en las actividades de preservación de las instituciones de ciencias naturales y biológicas.

En el jardín, generaciones de mariposas se desenvuelven a través de los ciclos del nacimiento, la vida y la muerte. El jardín no conserva mariposas individuales; funciona como un sistema que perpetúa las especies de mariposas. Ninguna de ellas es copia exacta ni clon de otra. Al contrario, la preservación ocurre en esta colección viva como consecuencia de la variabilidad de las mariposas individuales. Vale la pena señalar que éste no es el único caso. Lo mismo se aplica a todo tipo de colecciones vivas que se encuentran en zoológicos, acuarios, jardines botánicos, parques, reservas de vida silvestre y santuarios.

Los especímenes sujetos con alfileres en las cajas al otro extremo del pasillo se preservan de una manera diferente. En esta situación, primero se mataron las mariposas y luego se trataron y encajonaron de tal manera que sus restos se deteriorarían lo más lento posible. Esto es en sí un indicativo de toda una línea de prácticas de conservación para colecciones de especímenes de museos de historia natural, así como de colecciones de herbarios: se recogen muestras individuales y luego son tratadas de tal manera que sus restos pervivan en el futuro.

El jardín de mariposas opera en el marco folclórico, mientras que las mariposas preservadas en las cajas funcionan bajo la lógica del marco artefactual. El ADN de cada mariposa en el jardín puede considerarse como su identidad informativa. Aunque en cualquier mariposa existe más información que la que encontramos en su código genético, éste es, en términos generales, información codificada que se ha replicado. Si empezáramos a clonar mariposas, entonces estaríamos haciendo el proceso informativo, pero, cuando las mariposas se reproducen solas, la variabilidad de sus códigos genéticos es la clave para mantener una población determinada. Las mariposas preservadas en cajas funcionan de acuerdo con la lógica del marco artefactual. Cada una es históricamente contigua con otra que murió y fue inmovilizada hace años. Bien

²⁵ Conozco estos ejemplos particularmente bien porque mi esposa y yo nos casamos frente a ese mismo jardín de mariposas.



puede ser que algunas especies se hayan extinguido y que estos restos sean lo único que nos brinde acceso a la información sobre ellas. En este marco artefactual, las mariposas inmovilizadas son parte de los enfoques de la preservación tanto histórica como de la ciencia de la conservación.

Este no es el único caso de cómo se conservan los registros de la vida en las instituciones culturales. Es evidente que los procesos son similares a la manera en que se preserva el propio mundo natural. Por ejemplo, los fósiles operan bajo la lógica de los artefactos. A través de una variedad de circunstancias, se conservan los restos de algunas criaturas muertas, y al ser descubiertas se pueden estudiar para comprender su vida en el pasado. Los fósiles son históricamente contiguos a las criaturas que alguna vez vivieron, a pesar de que son muy diferentes de aquéllas en el momento en que estaban vivas. Al mismo tiempo, los biólogos evolucionistas también hacen uso de “fósiles vivientes”, como el ornitorrinco, los cuales son, en términos funcionales, reliquias vivas que demuestran las características de los animales del pasado, pero que se han preservado porque un ecosistema en particular no creó presiones selectivas que los eliminara. Estos especialistas usan la combinación de los fósiles reales y los fósiles vivientes, pues ambos conservan información sobre la vida en el pasado, aunque lo hacen a través de mecanismos en principio distintos, y la toman como base para reconstruir el desarrollo de las especies a lo largo del tiempo.

Llevar los marcos hacia adelante

He sugerido tres formas diferentes de enmarcar el trabajo de preservación. Al hacerlo, he usado una variedad de ejemplos para enfatizar las formas en que cada uno de ellos funciona, bajo lógicas muy diferentes que, a pesar de ser consistentes a nivel interno, a menudo entran en desacuerdo entre sí. Cada marco es distinto y, de alguna manera, incompatible. Sin embargo, en la práctica, los marcos también se cruzan e interactúan al realizar el trabajo de preservar cosas.

En el marco artefactual, intentamos extender la vida de los medios físicos. La continuidad histórica del artefacto es el foco de la preservación. En el marco informativo, trabajamos para establecer con claridad los criterios para copiar información codificada de un medio a otro. En este caso, el medio físico sólo es un portador o anfitrión de la información codificada. En el marco folclórico, la variabilidad y la hibridación de la información juegan un papel clave en la forma en que las historias y las secuencias se preservan, pero también cambian y se adaptan a las nuevas circunstancias. Aquí los enfoques de la documentación que ilustran esa variación son parte central del enfoque de preservación, pero, al mismo tiempo, los ecosistemas culturales o



biológicos mismos sirven como sistemas de preservación, pues transmiten la información que es más útil para sus necesidades contemporáneas.

Todo esto sirve para establecer un conjunto de marcos de referencia que podemos tener en cuenta cuando pensamos en los objetos digitales. Con esto en mente, como se ha demostrado en varios puntos, son las posibilidades formales y materiales de los medios de los objetos las que han jugado un papel clave en establecer lo que significa la preservación en diferentes contextos a lo largo del tiempo. Con ese fin, es fundamental que cualquier persona interesada en la preservación digital comprenda algunos de los aspectos clave de las formas y los medios digitales, que es el tema del siguiente capítulo.

Capítulo 2

La importancia de entender los objetos digitales

La preservación digital requiere de un entendimiento base de la estructura y la naturaleza de la información y los medios digitales. A lo largo de este capítulo intentaré proporcionar dichos antecedentes a través de algunos puntos clave que surgen de tres líneas académicas del estudio de nuevos medios que están relacionadas. Primera: toda la información digital es material. Segunda: las bases de datos son una forma mediática esencial para entender la lógica de los sistemas de información digital. Tercera: la información digital se entiende mejor como parte de un conjunto de plataformas anidadas. Comenzaré dando un resumen de estos tres puntos y luego ahondaré más en ellos.

Siempre que la información digital es almacenada, se encuentra físicamente codificada en un medio. Es fundamental reconocer que el flujo de bits en bruto (la secuencia de unos y ceros codificada en el medio original) tiene una materialidad tangible y objetiva. Esto representa un principio esencial de la preservación digital. Sin importar lo complejo que sea un sistema digital, en algún lugar, guarda información binaria codificada en un medio físico y tangible. Por lo tanto, es posible establecer cuál es la secuencia de bits completa de un medio en particular, o de un archivo específico, y crear una especie de huella digital que después puede ser utilizada para verificar y autenticar copias perfectas. Hay una inherente linealidad de la información digital en este nivel físico del flujo de bits. Donde sea que esté codificada la información, hay un primer y un segundo y un último bit. Sin embargo, esta característica es abstraída en la recuperación y el manejo de la información, lo que da como resultado algo muy distinto a la mayoría de nuestros otros medios.

La lógica de los medios computacionales es, en gran medida, la misma de las bases de datos. Así como el índice o el código pueden proporcionar una



metáfora valiosa para el orden y la estructura de un libro, la investigación y los estudios sobre nuevos medios sugieren que las bases de datos son y deben ser abordadas como la metáfora fundamental de los medios digitales. Desde esta perspectiva, en una base de datos no existe una “primera fila” constante, en su lugar, la presentación y orden de la información digital se basa en la consulta o búsqueda específica realizada sobre los datos. Dado que las bibliotecas y archivos han basado por largo tiempo sus concepciones de orden en las propiedades de los libros y los documentos en papel, la adopción de esta nueva lógica tiene implicaciones importantes para permitir el acceso a los materiales digitales a largo plazo.

Los sistemas digitales son plataformas puestas una sobre la otra. Si bien los flujos de bits son lineales, la lógica de las bases de datos en los nuevos medios funciona de una forma muy diferente. Al nivel de la base, esos flujos de bit lineales son animados, reproducidos, manipulados, alterados y puestos en uso e interacción a través de capas anidadas de plataformas. Al acceder a un objeto digital, los dispositivos computacionales interactúan con las estructuras de los sistemas de archivo, con los formatos de archivo y con otras varias capas adicionales de *software*, protocolos y *drivers*. Explicaré un par de ejemplos que ayudarán a aclarar este concepto. La experiencia del rendimiento de un objeto en una pantalla en particular, como jugar un videojuego o leer un documento, puede en sí misma ofuscar muchos de los aspectos importantes de un ese objeto, como las reglas de un videojuego o el texto eliminado de un documento que todavía existe pero no se muestra en la pantalla.

Como resultado de esta naturaleza de plataformas anidadas, los límites de los objetos digitales con frecuencia son completamente dependientes de la capa que se considera como la más significativa para un propósito específico. En este contexto, la forma y el formato digital deben ser comprendidos como un tipo de contenido. A través de las capas de la plataforma, los objetos digitales siempre son una multiplicidad de cosas. Por ejemplo, un videojuego Atari es un objeto tangible que usted puede tomar; es también una secuencia binaria de información, codificada en un medio físico, que es idéntica a todas las copias del juego; es un código fuente que tiene autoría como obra creativa; es un producto empacado, vendido y promocionado para una audiencia, y es además el símbolo de una variedad de aspectos de un momento histórico particular.²⁶ Cada uno de estos objetos pueden coexistir en las capas de la plataforma de ese

²⁶ Para una explicación elaborada de las cosas que representa un juego Atari, ver Ian Bogost, *Alien Phenomenology, or, What It's Like to Be a Thing*, Posthumanities 20 (Mineápolis: University of Minnesota Press, 2012).



objeto tangible, pero se debería desarrollar un enfoque diferente de preservación dependiendo de cuál sea la capa significativa para un fin concreto.

En los párrafos siguientes, ahondaré en cada uno de estos tres puntos: la materialidad de los objetos digitales, la lógica de las bases de datos y la naturaleza de la plataforma de los nuevos medios. También mencionaré algunos ejemplos con el objetivo de demostrar que estos puntos son relevantes para entender los objetos digitales de una forma que nos permita comenzar a pensar en cómo preservarlos. Luego explicaré algunos ejemplos que unen estos conceptos con otros puntos conceptuales clave para la comprensión de los medios digitales.

Lo digital siempre es material

Gran parte del lenguaje que utilizamos para hablar sobre medios digitales ensombrece la materialidad de la información digital. Mis archivos de Google Docs están ahí en algún lugar en la metafísicamente llamada “nube”. No hay que equivocarse: la nube se compone de computadoras de otras personas. Esto es un desafío central en el trabajo con objetos y medios digitales. El disco duro en su computadora a veces es una caja negra en todo el sentido de la palabra. Dentro de esa caja hay un mecanismo complejo que codifica información a una escala tan pequeña que no es visible.²⁷

Hasta cierto punto, la mayoría de la gente entiende que la información digital está compuesta de bits, lo que a veces invoca una imagen de unos y ceros circulando en una pantalla, como en la película *The Matrix*. Si usted busca una imagen en Google para el término “bits”, encontrará alguna ilustración con unos y ceros en un monitor, varias de las cuales han sido tomadas desde un ángulo que sugiere acción y movimiento. Lo que no encontrará son imágenes como la de abajo. Esos unos y ceros son la unidad básica más pequeña de la información digital. Sin embargo, esa forma de visualizar estos elementos no logra capturar la naturaleza material y física de la información digital.

Aquí usted puede ver las magnetizaciones de bits individuales codificados en bandas sobre la placa de un disco duro. A este nivel, el de los micrones, cada unidad de información inscrita es a todas luces tangible y física. Cada bit tiene dimensiones. En el fondo, la lectura y escritura en el sustrato de cualquier medio subyacente implica una traducción de lo digital a lo análogo y luego en reversa. Es decir, un bit de información es codificado en el medio y después es releído.

²⁷ El argumento de la materialidad de los objetos digitales es una versión corta de muchos de los temas de los dos primeros capítulos de Kirshenbaum.

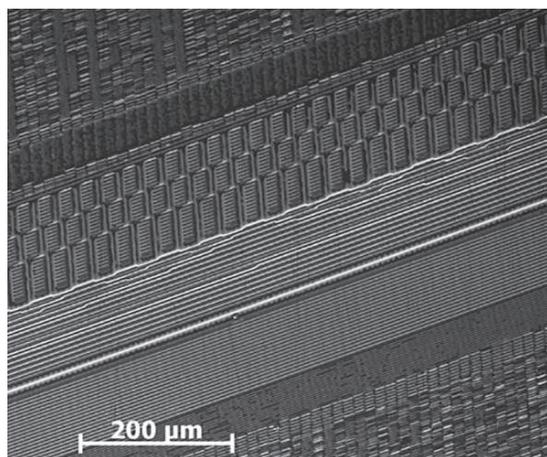


Figura 1. Grabación de magnetizaciones individuales de bits en una placa de disco duro de 200 MB (grabación visualizada utilizando CMOS-MagView).

Tomado de Wikimedia Commons, subido por Matesy GmbH.

Lo anterior no sólo es cierto para los discos duros. Usted, de hecho, puede ver los bits individuales en una tarjeta o en un rollo de cinta perforados. La presencia o ausencia de una perforación es leída como un “uno” o un “cero”, y es fácil notar que cada una tiene un largo y un ancho. Algunas pueden ser un poco más anchas o más largas que otras, o estar un tanto desalineadas, pero, mientras estén lo suficientemente ajustadas para poder ser leídas e interpretadas, funcionan para almacenar la información codificada. Un CD regrabable usa un láser para codificar bits sobre una capa de color. En la mayoría de los CD o DVD comerciales, la información está codificada en hendiduras pequeñas llamadas *pits* o “fosas”. Éstas son leídas por un láser en un reproductor de CD o DVD. En el caso de las memorias *flash*, que contienen las USB que muchas personas traen consigo en sus llaveros y que son también la base de las memorias dentro de los teléfonos móviles, los electrones son guardados en algunas partes del medio. Esos electrones existen en el límite de lo que podemos considerar material. Sin embargo, tenga por seguro que los electrones son, sin duda, parte de nuestro universo material y que no existen en ningún tipo de plano metafísico.

Estos ejemplos sirven para subrayar un punto fundamental. Toda la información digital es material. Esto tiene una consecuencia crucial en el trabajo de la preservación digital. No importa lo complejo o desafiante que sea un objeto u obra digital, en algún lugar hay un objeto físico con una secuencia de



marcas de forma lineal o un flujo de bits. Esa secuencia de bits se puede leer desde el mismo medio o copiar a otro, y su precisión puede comprobarse. Este proceso de “preservación de bits” es más o menos directo y fácil de verificar. Sin embargo, esos bits sólo son en verdad útiles para nosotros cuando podemos reproducirlos e interactuar con ellos, lo que es posible a través de otro de los aspectos clave de los medios digitales: las bases de datos.

Cualquier cosa puede ir primero en una base de datos

Clic derecho. Ordenar resultados. Ésta es una de las operaciones más comunes que los usuarios de computadoras utilizan en todo el mundo. En muchas aplicaciones es posible ordenar también canciones, emails, etc., por nombre, por fecha de creación o por la última vez que se abrieron, así como por autor, tipo o tamaño de archivo. En la mayor parte de los sistemas operativos también se puede cambiar la información que se quiere ver sobre cada uno de esos archivos. Es posible verlos como una lista, como íconos grandes o pequeños, o con todos sus detalles en columnas, cuyos encabezados también permiten ordenar la manera en la que se muestra la información. La gramática de estas interfaces, filtrar y ordenar, se reproduce en casi todos los niveles de abstracción en la computación. El investigador de medios Lev Manovich explica sobre este punto que, en cuanto “forma cultural, las bases de datos representan el mundo como una lista de ítems que se rehúsa a ser ordenada”.²⁸ Nuestra relación con la información digital está arraigada en la base de datos a través de la interfaz: es posible interactuar con el orden y la secuencia subyacentes del flujo de bits mediante una multiplicidad de órdenes. Es reemplazada por una variedad de posibilidades de secuencia. Esto es fundamentalmente diferente a la forma en que interactuamos con los medios análogos.²⁹

Es verdad. Incluso al nivel básico del disco duro que hemos explorado, los discos están divididos en sectores. Éstos son administrados a través de tablas. Si bien la información en el disco, el flujo de bits, está ordenada en una forma lineal, éste abstrae sus sectores en un volumen coherente de espacio que se pone a disposición de la computadora para que usted lo utilice para almacenar archivos.

La lógica de las bases de datos se entiende mejor si se contrasta con parte de la lógica lineal y referencial que ha sido parte clave de otros tipos de medios.

28 Lev Manovich, “Database as a Genre of New Media”, *AI & Society* 14, núm. 2 (2000): 176-183, DOI: <https://doi.org/10.1007/BF01205448>.

29 Manovich habla más sobre el enfoque de las bases de datos en *The Language of New Media* (Cambridge, Massachusetts: MIT Press, 2002).



Por ejemplo, el códice, es decir, un manuscrito encuadernado, me permite hacer referencia a la página 23 de este libro y le permite a usted ir rápido a ella y ver lo que he referenciado. De la misma forma, usted puede ir al índice en la parte final de un libro y ver un orden distinto del lineal que se produce al voltear cada página individualmente. Si le interesara ver dónde se habla sobre bases de datos, podría buscar la entrada para “bases de datos” en el índice e identificar cada página en la que se menciona el concepto en el libro.

Los índices de los libros tienen algo de esa lógica de las bases de datos. Es un poco como elegir maneras distintas de ordenar un libro. Es algo similar a la manera en que usted ordenaría la información en su computadora. Más allá del índice, hay formas más tardadas de ordenar el contenido de un libro, como las concordancias. En ellas, se presenta una lista de palabras clave de un libro junto con su contexto. Entonces, si este libro tuviera una concordancia, incluiría cada oración en donde aparece la palabra, lo que le permitiría leer el contexto en un solo lugar en vez de ir a cada página para verlo.

La lógica de las bases de datos ha estado con nosotros por mucho tiempo. No se trata de algo del todo nuevo que haya aparecido con los medios digitales. Sin embargo, no se utilizaba predeterminadamente. Costó mucho trabajo lograr que los medios analógicos actuaran como bases de datos. Por defecto, un libro se lee dando vuelta a las páginas una por una. En contraste, al leer el mismo libro en una computadora o dispositivo móvil es muy probable que primero el usuario presione Ctrl+F y busque el término que más le interesa y use eso como punto de partida en el texto.

La búsqueda es en sí uno de esos aspectos centrales de la lógica de las bases de datos. Nosotros no leemos bases de datos. Realizamos búsquedas en ellas. Uno de los mitos clave de la World Wide Web (sin decir que es falso, pero sí que ha alcanzado ya un nivel de repetición que le ha dado la función de leyenda) es que Yahoo, que comenzó como “Jerry and David’s Guide to the World Wide Web”, funcionaba a partir de la lógica de “organizar contenido”, es decir, construyendo a mano índices de sitios basados en temas. En contraste, los advenedizos en Google desarrollaron PageRank, un algoritmo que indexaba computacionalmente la red y sopesaba los resultados con base en los vínculos entre páginas web. Puede ser útil volver a repensar esta historia para los bibliotecarios y archivistas. ¿Estamos tratando de parecernos a Yahoo y manejar el contenido digital como si creáramos el índice de un libro, en el que pasamos tiempo considerable intentando organizarlo y describirlo, en vez de actuar como Google y utilizar nuestro tiempo buscando una forma de sacar a la superficie las relaciones del análisis de datos subyacentes?



Otro aspecto fundamental es que, mientras el orden y la secuencia han sido parte central de la administración de las colecciones en bibliotecas, la lógica de las bases de datos hace que éstos pierdan importancia o, más específicamente, los medios computacionales permiten una multiplicidad de órdenes y secuencias. Un sistema de archivos vertical o de libros en estanterías organizados de acuerdo con el sistema de clasificación de la Library of Congress tiene un orden basado en los límites tangibles del espacio físico. Cada cosa puede estar en un solo lugar. Así, en muchas universidades usted puede caminar por las estanterías de la biblioteca y encontrar este libro, posiblemente alrededor de la letra “z”, con otros libros sobre ciencias de la información. Sin embargo, no podrá hacer clic derecho en la estantería y ver todos los títulos en orden alfabético por autor o cualquier otro encabezado temático, cosa que sí puede usted hacer en los registros en línea del catálogo, pero, en este caso, son los objetos mismos los que se cambian de lugar y se reorganizan. Este tipo de orden ha formado parte importante del trabajo de bibliotecas y archivos, y tiene sentido que así sea, pues ha sido clave en nuestra habilidad para encontrar y tener acceso a estos objetos a través de sus ubicaciones físicas. De manera relevante, el desarrollo del catálogo de fichas es, en sí, un ejemplo claro de cómo las bibliotecas han utilizado por mucho tiempo la lógica de las bases de datos para permitir un número variable de órdenes.³⁰ Es un tema que trataré más adelante en el libro. Por ahora, vale la pena destacar que desde hace mucho la lógica inherente de las bases de datos ha sido parte del trabajo de las bibliotecas y de las ciencias de la información y que siempre ha sido un esfuerzo gigante, algo que funcionaba contra corriente en los materiales como libros, papel y registros. En contraste, los medios digitales usan la lógica de las bases de datos por defecto. Esto es algo en lo que nuestro campo tiene todavía un largo camino por recorrer para ser aceptado completamente.

De aquí en adelante, es importantísimo recordar que, primero, siempre habrá una secuencia de unos y ceros codificados en un medio tangible y que, segundo, a través de todos los niveles de abstracción que vamos a discutir, esa información es accesible y manipulable mediante la lógica de las bases de datos. El resto de este capítulo se construye a partir de estos dos puntos, comenzando por el medio tangible y hasta la exploración de las relaciones entre los distintos tipos de plataformas en los objetos digitales. En este contexto, después

30 Para una historia completa del desarrollo del catálogo de fichas ver Markus Krajewski, *Paper Machines. About Cards & Catalogs, 1548-1929*, History and Foundations of Information Science (Cambridge, Massachusetts: MIT Press, 2011).



desglosaré una serie de características fundamentales de los medios digitales que provienen de la naturaleza de las plataformas.

Se trata de plataformas a lo largo y ancho

Todas nuestras interacciones con la información digital tienen como intermediario las capas de las plataformas. Siguiendo el trabajo realizado en el campo del estudio de las plataformas, he decidido utilizar una definición más bien amplia: cualquier cosa que el programador dé por supuesto cuando desarrolla, y lo que sea que, del otro lado, el usuario requiera que esté funcionando para poder utilizar un *software* en particular, eso es la plataforma. Bajo el apelativo “estudio de las plataformas”, ha surgido un cuerpo entero de investigación académica que busca explorar la manera en que las plataformas computacionales “permiten, limitan, dan forma y apoyan el trabajo creativo que se hace en ellas”.³¹ Lo anterior incluye, pero no se limita, a los sistemas operativos, los lenguajes de programación, los formatos de archivo, las aplicaciones de *software* para crear o reproducir contenido, los esquemas de codificación, los algoritmos de compresión y los protocolos de intercambio.

La mejor forma de explicar la lógica de las plataformas es a través de un ejemplo. Podemos comenzar con un caso de archivos perdidos, tomado de mi historia personal. Una copia del sitio web para la banda que comencé en la preparatoria. Por desgracia, los archivos se perdieron, pero la relativa falta de sofisticación los hace un buen caso de estudio para analizar.

Imagine que le entrego un disco duro externo con una copia del sitio web de mi banda. Así que lo conecta a su computadora y lo abre para explorar el contenido. En el directorio de origen ve un archivo llamado “index.html”, más otras carpetas. Hace clic en el archivo y su navegador web lo inicia. Puede ver ahí mismo una representación de la página. En la página de inicio se muestran algunas fotos de la banda; aparecen simplemente en su navegador. Usted puede dar clic y leer la biografía de cada uno de los miembros en otra página HTML. Incluso puede hacer clic en una página más para acceder a una de las canciones en formato MP3. Si hace clic nuevamente, se reproduce en su navegador. Así es como usted puede disfrutar o, al menos, experimentar la música de Marzapan (*ca.* 2001).

Ésta sería sólo una forma de experimentar esos archivos. Su navegador web es una plataforma para ver sitios web y, como tal, es un medio en el que

31 Nick Montfort e Ian Bogost, *Racing the Beam. The Atari Video Computer System*, Platform Studies (Cambridge, Massachusetts: MIT Press, 2009), 2.



puede interactuar con este sitio, tal y como el creador tenía pensado para los usuarios. Pero, si cierra el navegador y comienza a abrir las otras carpetas en ese directorio, podrá examinar cada uno de los archivos individuales que conforman el sitio.

Al igual que con cualquier otra carpeta de archivos, usted puede organizarlos por tipo, nombre, fecha, etc. Mientras explora los directorios es posible que aprenda algunas cosas nuevas sobre el sitio web de la banda. Hay elementos que no eran visibles en la pantalla de su navegador. Desde este punto, puede observar que algunos archivos fueron creados en 2001 y otros, en 2003. Puede hacer clic en el mismo archivo “index.html” y abrirlo en una aplicación distinta, como un editor de texto, y ver todo el código HTML que fue interpretado por su navegador para crear la página. Puede incluso haber comentarios en el HTML con información adicional sobre por qué se presenta de esa manera.

Si hace clic derecho en uno de los archivos MP3, debería iniciarse cualquiera de las aplicaciones que su computadora tenga predeterminadas para reproducir archivos de audio. También podría hacer clic derecho en el archivo y seleccionar “ver las propiedades del archivo”. En ese caso, es posible ver los metadatos que están incrustados en el archivo: que el autor del archivo se llama “Marzapan” y que al parecer es de un álbum titulado *Beauty of a Bad Situation*. Toda esa información está incrustada dentro de las etiquetas ID3 del archivo.

Si usted quiere ver cómo funciona, sólo cambie la extensión del archivo de “.mp3” a “.txt” e ignore el mensaje algo pasivo-agresivo de la computadora que le pregunta si “está seguro de que quiere hacer eso”. Ahora, al hacer clic en él, se abrirá el programa que usted utiliza para los archivos de texto y se mostrarán un montón de caracteres sin sentido, pero también verá todos esos datos que aparecieron en las propiedades del archivo, es decir, el nombre del álbum y del autor. Estos metadatos incrustados están codificados como texto dentro del archivo, de modo que puede verlos con cualquier herramienta que pueda leer ese tipo de texto incrustado que, en este caso, es Unicode. Hay una gran posibilidad de que pueda observar texto adicional aquí también, pues todo lo que era visible en las propiedades del archivo, a través del sistema de archivos, sólo se trataba de la información que los creadores de dicho sistema pensaron que era particularmente relevante.

Si continúa haciendo clics, encontrará que, aunque sólo pudo ver uno de los archivos MP3 mediante el navegador web, de hecho ¡el EP entero, con sus cinco canciones, está ahí! Esta clase de patrón se repite a través de todas estas capas de plataformas. A medida que continúa explorando más a fondo en la estructura, descubrirá cosas que no eran visibles al nivel más alto de abstrac-



ción, desde la forma en que se reproduce hasta la estructura en el sistema de archivos y la información incrustada en ellos. Hasta ahora, ha logrado ver el sitio web como es reproducido, los archivos individuales y sus directorios, y también algo de la información incrustada en algunos de esos archivos individuales. Si quisiera, usted podría hacer una copia lógica de todos los datos en su propio escritorio y aun así es posible que haya más información en ese disco que todavía no ha sido mostrada.

En otro nivel, usted podría abrir cualquiera de los archivos que encuentre en el disco en un editor Hex, una aplicación que le permite ver los datos binarios subyacentes de un archivo. Al pasar cualquier archivo por este proceso, usted aprendería bastante sobre la secuencia y la estructura real de la información que contiene.

Si lo desea, usted incluso podría crear una copia forense de toda la información en el disco. Cuando la versión más joven de mi persona eliminó cosas del disco, creía que en verdad estaba borrando la información, pero no era cierto. Como es común con muchos medios de almacenamiento digital, al borrar algo, en general sólo estamos diciéndole al sistema que olvide que hay información ahí y que marque ese espacio como disponible para ser reutilizado. Entonces, si creara una copia forense de la secuencia de bits tal y como está distribuida en el disco, y luego utilizara otra vez un editor Hex para analizar lo que realmente está almacenado en cada área del disco, es muy posible que encuentre que gran parte del espacio que se presumía vacío todavía tiene archivos que se podrían recuperar.

Dependiendo de cómo fue borrado exactamente, es posible que necesite enviarlo a un grupo forense especial que pueda ignorar lo que dice el disco sobre la información en él y forzar la lectura de las partes que se cree están en blanco. Aunque es posible acceder a la información que ya fue sobrescrita, usted bien podría encontrar que, antes de poner el sitio web de mi banda en el disco, éste contenía todos los trabajos que escribí en la escuela en archivos .doc y podría abrir cualquiera que aún se encontrara ahí.

Ése fue un *tour* rápido por las capas de las plataformas múltiples, que están apiladas una sobre otra, en el caso de este disco duro. La tabla de abajo muestra todo esto como una serie de capas.³²

³² La tabla y gran parte del análisis están basados en Christopher Lee, “Digital Curation as Communication Mediation”, en *Handbook of Technical Communication*, ed. de Alexander Mehler y Laurent Romary (Berlín; Boston, Massachusetts: De Gruyter Mouton, 2012).



Nivel	Explicación	Ejemplo
Objeto compuesto o complejo	Un objeto compuesto por múltiples archivos, pero percibido como una unidad	El sitio web como es reproducido en un navegador
Archivo reproducido	Un archivo reproducido a través de una aplicación de <i>software</i>	Los archivos MP3 o JPEG del disco vistos a través de reproductores o visores
Archivo en el sistema de archivos	La información sobre archivos individuales visible a través de un sistema de archivos	Mirar los directorios de archivos en un disco, para ver sus propiedades individuales
Archivo como flujo de bits	La secuencia lineal de valores binarios en el archivo	Abrir el archivo en un editor Hex
Información de subarchivo	Información extraíble y visible del interior del archivo	El texto en las etiquetas ID3 incrustados dentro del MP3 visible en la versión de texto
Flujo de bits a través de 1/0	Serie de unos y ceros como son presentados a la computadora	Los contenidos del disco que se pueden copiar
Flujo de bits en un medio físico	Codificación física de la información interpretada en el medio subyacente	Las características físicas del disco, en este caso las magnetizaciones sobre el disco, que pueden aún contener información adicional

Al recorrer la tabla desde abajo hasta arriba, se aíslan las capas individuales con sus distintas posibilidades y propiedades. Es importante destacar que la capa de la base, la del flujo de bits en un medio físico, es la única en verdad material. Volviendo a los conceptos del capítulo anterior, es en este nivel que, potencialmente, interactuamos con un objeto o artefacto. Todo lo que está por encima es informativo, es decir, al moverse de la primera capa a la segunda, se pasa de la información codificada en un medio a las señales digitales binarias.

Piense en las capas de esta pila como el objeto en el que se enfoca la preservación. Además, cada una representa diferentes conceptos de un objeto que los usuarios presentes o futuros podrían estar interesados en estudiar. Por ejemplo, para alguien que estudia sitios web publicados, será fundamental enfocarse en la capa superior. En contraste, para aquellos interesados en un disco como éste, que podría ser parte de una colección de manuscritos, los aspectos relacionados al cómo se organizó la información e incluso los archivos eliminados pueden



ser de interés. En resumen, identificar qué capas de la plataforma son significativas para los propósitos de una organización se vuelve una parte vital de la práctica de la preservación digital.

La naturaleza de las plataformas de los nuevos medios representa una oportunidad para ahondar en una serie de puntos conceptuales clave, que abordaré utilizando el ejemplo del sitio web de Marzapan: el *screen essentialism*, es decir, esencialismo de pantalla, o la teoría de formato, la compresión, los protocolos de intercambio, los esquemas de codificación y el código y los archivos fuente. Estos conceptos proporcionan una vía para la comprensión de un determinado conjunto de contenido digital para propósitos de acceso y de preservación.

Evitar el esencialismo de pantalla

Al inicio del ejemplo anterior miramos el contenido del sitio web a través de un navegador. Si preguntara qué es ese sitio, es probable que la mayoría de la gente contestara que se trata precisamente de eso, lo que está siendo mostrado en la pantalla, y hasta cierto punto tendrían razón. El sitio web es lo que se ve. Sin embargo, como descubrimos al indagar con más profundidad en los archivos, existe mucha más información que puede ser interesante en las distintas capas de estas plataformas. Hay una tendencia cuando se trabaja con la información digital a dar por hecho que la manera en que luce algo en la pantalla es lo más significativo. Como muestra el ejemplo anterior, esta idea puede ser muy engañosa. Los investigadores de los nuevos medios lo han llamado “esencialismo de pantalla”, un término un tanto burlón.³³ Afirmer que alguien es un existencialista de la pantalla es decir que no está prestando atención a todos los otros aspectos de ciertos objetos que podrían ser significativos e importantes, pero que no aparecen en ella.

Como ilustra el ejemplo anterior, en casi todas las capas anidadas en la pila de plataformas hay información que puede ser significativa y que no siempre aparece en la experiencia común de interactuar con el archivo en su aplicación predeterminada. Desde los sectores del medio físico hasta los archivos mostrados en el sistema de archivos, como también en los metadatos incrustados en el archivo mismo, existen todos estos rincones y rendijas en los que la información que podría ser interesante está esperando a ser interpretada. La forma de evitar este problema es dando un paso atrás para asegurarnos de que tenemos

³³ El término *screen essentialism* fue acuñado por Nick Montfort en “Continuous Paper: MLA” (conferencia, MLA, 28 de diciembre de 2004), http://nickm.com/writing/essays/continuous_per_mla.html.



una comprensión sólida de todas las capas de las plataformas que están en juego en un contexto específico y luego cerciorarnos de que hemos considerado con cuidado cuáles, en particular, son relevantes para lo que deseamos preservar sobre un objeto.

Comprender la teoría de formatos

Mientras explorábamos el disco duro, nos encontramos repetidamente con formatos de archivo. A ese nivel, las extensiones (.mp3, .jpg, .doc, etc.) se vuelven actores clave en nuestra historia. El formato de archivo y la aplicación reproductora son cosas que nos permiten interactuar con el objeto digital. En este momento, podemos comenzar a pensar en movernos de forma horizontal en la tabla presentada anteriormente. La pila vertical de capas digitales, detalladas ahí, describe de forma precisa qué es la información en el disco, pero no cómo deben ser invocadas cada una de las capas de la plataforma para interactuar con el contenido. Los formatos de archivo son un buen ejemplo de este tipo de conexión horizontal.

Los formatos de archivo hacen posible la computación moderna. Son la convención que establece las reglas sobre cómo se estructura y almacena la información en un archivo. Sus extensiones (.mp3, .jpg, .doc), en parte, lo definen, así como definen la manera en la que puede ser interpretado. Éstas también le dicen al sistema operativo de una computadora qué aplicación debe usar para reproducirlo. Usted también puede identificar archivos basándose en la información de su encabezado o su firma. Cada formato de archivo tiene su propia historia. Hasta cierto punto, conceptualiza un aspecto del mundo en el que se desenvuelve. Por ejemplo, el éxito y la diseminación del uso del PDF reside en qué tan bien replica el aspecto de los documentos impresos. Los PDF “comparten la forma y fijeza³⁴ de las impresiones que otros formatos no comparten”.³⁵ En ese sentido, este formato de archivo se entiende mejor como parte de la historia de compartir y reproducir documentos que se remonta a las tecnologías como la xerografía y el microfilm. En relación con esto, es importante entender que el formato de un archivo es más que un contenedor de información, también es informativo en sí mismo.

34 N. del T.: en inglés el término es *fixity* y en esta obra se traduce en algunos casos como “integridad de los datos” y en otros como “fijeza”, según el contexto en que se encuentra.

35 Lisa Gitelman, *Paper Knowledge. Toward a Media History of Documents* (Durham; Londres: Duke University Press, 2014), 114.



Los archivos MP3 ofrecen otro ejemplo. Como muchos formatos, éste utiliza la compresión. Un archivo comprimido, en términos generales, es más pequeño en cuanto a su tamaño que otros que codifican de manera secuencial toda su información. Como contraste, es normal que un archivo .wav contenga datos de audio de “modulación por impulsos codificados” (PCM) no comprimidos. Esos datos incluyen, en orden lineal, información secuencial para cada estado del sonido, por lo que, cerca del inicio del archivo, usted encontrará partes relacionadas con el comienzo de la grabación y, hacia el final, bits que son parte del sonido. Explorar un poco la compresión ayudará a establecer lo que podría ser significativo sobre un archivo en un contexto particular.

Entender la compresión

Un archivo comprimido es más pequeño que uno no comprimido. Suena como si alguien lo hubiera puesto en una prensa o algo así, pero en computación la compresión se logra a través de la remoción sistemática de información. En el caso de la compresión sin pérdida, el archivo sólo removerá la información que es redundante o idéntica. Por ejemplo, un archivo de imagen, en vez de almacenar valores para “pixel rojo, pixel rojo, pixel rojo”, puede codificar la secuencia como “tres píxeles rojos”. También existe la compresión con pérdida, en la que se toma una decisión sobre qué información en particular no es relevante, y luego dicha información se remueve de forma sistemática. La compresión con pérdida hace posible reducir de forma dramática el tamaño de un archivo.

En el caso de un MP3, pueden utilizarse diferentes niveles de compresión para crear archivos cada vez más pequeños. Es probable que un archivo que está ligeramente comprimido no suene muy distinto para usted. Esto es porque los algoritmos utilizados para comprimir archivos de audio se fundamentan en casi una centena de investigaciones sobre procesamiento de señal. Las compañías telefónicas gastaron recursos considerables para generar métodos que identifiquen qué partes de la información de audio están sintonizadas con la audición humana. Esos experimentos identificaron personas que se consideraba que tenían un “oído de oro” y entrenaron herramientas y sistemas basadas en sus gustos y sensibilidades. En este aspecto, cada MP3 “trae consigo entendimientos prácticos y filosóficos sobre lo que significa comunicar, lo que significa escuchar y hablar, cómo funciona el oído de la mente y lo que significa crear música”.³⁶ En el caso del MP3, esta historia también se apoya en

36 Jonathan Sterne, *MP3. The Meaning of a Format. Sign, Storage, Transmission* (Durham: Duke University Press, 2012), 2.



la ley y el mercado. Como formato patentado, cada aplicación que reproduce un MP3 debe pagar para poder hacerlo.

Encontrar archivos MP3 en un medio digital es relevante para la historia y trayectoria del formato. También comunica algo sobre el archivo de audio y sobre su historia y producción. Si, en vez de éstos, hubiéramos encontrado archivos WAV, que contienen datos no comprimidos, habría significado que al creador de los archivos le importaba más la calidad del sonido. Al mismo tiempo, seguramente causaría perplejidad, pues el tamaño de archivo más pequeño de un MP3 es parte de lo que lo hizo tan útil como formato para la distribución a través de Internet. Es decir, si hubiera habido archivos WAV en el sitio web, podría generar preguntas válidas sobre por qué se utilizó ese formato. De forma similar, si los archivos fueran OGG, un formato gratuito sin restricciones de patentes de *software*, este hecho podría comunicar que el creador del archivo estaba más comprometido con el movimiento de *software* de código abierto que con asegurarse de que la mayoría de los usuarios que visitaran el sitio pudieran reproducirlo sin complicaciones en sus navegadores. En cada uno de estos escenarios, el formato del archivo y la información sobre cómo está estructurado el archivo dentro de los parámetros de sus especificaciones contiene información que, en muchos casos, es esencial para determinar lo que es importante de él.

Los protocolos de intercambio

También vale la pena destacar que los formatos de archivo están relacionados con otros tipos de plataformas auxiliares. Las restricciones de ancho de banda para transferir archivos a través de Internet han jugado un rol importante en el establecimiento del éxito o el fracaso de ciertos formatos de archivo en el pasado. Este punto está relacionado tanto a la infraestructura material de la red (el cable de fibra óptica, los proveedores de servicio, las torres de telefonía, etc.) como a la infraestructura que existe en protocolos como HTTP, TCP/IP y el sistema de nombres de dominio, que gobiernan la forma en que funciona la World Wide Web y la manera en la que se transmite la información. Así como la caja negra de un disco duro oculta la materialidad de la información digital, también lo hace la retórica y el diseño de estos sistemas, protocolos y plataformas.

Los estudios de medios han demostrado continuamente que comprender cómo funciona cualquiera de estas plataformas también significa revelar para quién funciona. Incluso algo que parece banal, como las especificaciones técnicas de estos protocolos, “importan, tanto ontológica como políticamen-



te”.³⁷ Aunque los protocolos no están presentes en el ejemplo anterior del disco duro, de todas maneras, están implicados. El contenido fue diseñado para que se acceda a él a través de Internet y, como tal, protocolos como HTTP, TCP/IP y el Sistema de Nombres de Dominio, entre muchos otros, además de la infraestructura física de la red, juegan un rol clave en el cómo y el por qué se ha estructurado de esa forma.

La codificación de la información

La información dentro de los archivos se codifica a través de estándares que también se especifican y existen fuera de los archivos mismos. Así es como, mediante un editor de texto, en el ejemplo del disco de mi banda, fue posible leer el contenido del archivo de audio. El texto dentro de cualquier archivo se codifica de acuerdo a estándares que, de igual modo que el formato, transmiten información y contexto sobre la historia del objeto.

En ese sentido, el texto puede codificarse de muchas maneras, por ejemplo, el ASCII viene con un conjunto de posibilidades para producir un grupo particular de caracteres, mientras que las versiones de Unicode incluyen posibilidades diferentes para codificar más tipos de caracteres. De acuerdo con la tendencia general de acercarse a estos elementos como plataformas, es crítico guardar la información sobre el esquema de codificación utilizado para que los futuros usuarios puedan estar seguros de atribuir la razón de ser de algo a la fuente correcta. Dicho de otra manera, si en algún momento un esquema de codificación no pudiera representar marcas diacríticas en el texto, usted no querría que alguien en el futuro creyera que la falta de éstas se debe a un simple descuido del autor. Esto nos lleva a un punto más amplio. Cualquier cambio a los archivos, ya sea para normalizarlos o migrarlos a una versión más nueva de un formato, o para actualizarlos con codificaciones mejoradas, puede generar un impacto en la habilidad del usuario para contextualizarlo de forma precisa dentro de las múltiples plataformas en las que está involucrado en un momento particular del tiempo.

³⁷ Eugene Thacker, “Protocol Is as Protocol Does”, en *Protocol. How Control Exists after Decentralization*, ed. de Alexander R. Galloway (Cambridge, Massachusetts: MIT Press, 2006), xii; para saber más sobre la manera en que funcionan los protocolos subyacentes de Internet como parte de los sistemas de control y vigilancia, ver Wendy Hui Kyong Chun, *Control and Freedom. Power and Paranoia in the Age of Fiber Optics* (Cambridge, Massachusetts: MIT Press, 2005).



La reproducción de los objetos

Nuestras experiencias con la información digital, generalmente, están mediadas a través de un número de interfaces gráficas que reproducen la información en nuestras pantallas y, en muchos casos, nos permiten interactuar con ella. Esto es, en parte, la razón por la que es tan fácil ser un esencialista de la pantalla. La experiencia de utilizar computadoras es la de interactuar con el contenido a través de interfaces en una pantalla. A cierto nivel, incluso algo como su sistema operativo reproduce información digital, y es la interfaz gráfica del usuario la que usted utiliza para navegar y manipular archivos. Cabe destacar que diferentes versiones de aplicaciones de *software*, o diferentes implementaciones para la lectura de tipos particulares de archivos, pueden tener como resultado que éstos se reproduzcan de formas muy distintas.

Como ejemplo, muchas obras tempranas del Net Art, creadas para ser desplegadas en la World Wide Web, dependían de peculiaridades y características que sólo estaban disponibles en los primeros navegadores, por lo que, en algunos casos, para capturar como se pretendía que fuera visto, es necesario grabar su apariencia en una pantalla utilizando el *software* y el *hardware* de reproducción considerados originalmente.³⁸

Los códigos y archivos fuente

Existe otro género completo de tipos de información que hacen posible gran parte de nuestra experiencia con los archivos y de cómo se reproducen. En el caso de los archivos de música en un CD, existen todas las pistas originales de audio para cada instrumento grabado, que después fueron mezcladas y manipuladas en algún *software* como Pro Tools, Acid Pro o algo como el Garageband de Apple. Para un archivo de imagen podría ser algo como Photoshop o Illustrator, que vienen con su propio formato de archivo, .psd y .ai, respetivamente.

En general, los archivos fuente son mucho más exigentes que los archivos que se producen para ser distribuidos. Es decir, un archivo de Photoshop sólo necesita ser abierto por ese *software*; en cambio, un JPG que haya sido exportado desde Photoshop está estandarizado al punto de que varios tipos diferentes de aplicaciones pueden reproducirlo. Además, cada versión nueva de Photoshop

38 Ben Fino-Radin, “Rhizome Artbase: Preserving Born Digital Works of Art” (conferencia, Digital Preservation 2012, 24-26 de julio de 2012), https://digitalpreservation.gov/meetings/documents/ndiipp12/DigitalCulture_fino-radin_DP12.pdf.



viene con nuevas características, por lo que el formato debe cambiar. Al mismo tiempo, el archivo fuente se puede modificar una cantidad casi infinita de veces y, en consecuencia, es mucho más complejo que el JPG que puede exportar para ser difundido. En un archivo de Photoshop, se puede jugar con todo tipo de capas, retocar, hacer y deshacer diferentes aspectos durante el proceso de manipulación de una imagen. En el caso del archivo fuente de audio, siempre se puede volver y aumentar la reverberación del bombo o cambiar la forma en que las guitarras panean desde el canal izquierdo al derecho, o regresar y grabar otro conjunto de voces.

La relación entre los archivos que se pueden reproducir sin complicaciones y que vemos o escuchamos y los archivos fuente que los generan también aplica para el *software*. La mayor parte de nuestra experiencia al utilizar un *software* implica ejecutar una aplicación que ya ha sido compilada, es decir, se escribió en un lenguaje de programación de computadoras como C o BASIC, que dio forma a su código fuente, utilizando comandos de texto. Ese código fuente luego se compiló y se utilizó un programa de computación para transformarlo en código de máquinas, el cual, generalmente, ya no es legible como texto para una persona. En vez de esto, está optimizado para correr de manera rápida y eficaz en una computadora. En el caso de algo como Photoshop, existe el *software* que uno instala para abrir y editar archivos de Photoshop, pero también está el código fuente que se compiló para crear el *software*.

Ambos son objetos que podrían ser interesantes para su estudio. Cuando alguien utiliza Photoshop para editar imágenes, tanto la interfaz como su historial pueden privilegiar y sugerir formas particulares de trabajar las imágenes. Además, el código fuente es en sí un registro que puede servir para entender y explicar el cómo y el porqué este *software* sugiere ciertas formas de manipular las imágenes. La manera en que Photoshop hace esto ha sido un tema de investigación en el área de estudios de *software*.³⁹ Incluso el Computer History Museum adquirió, preservó y publicó el código fuente del primer lanzamiento de Photoshop.

El desorden de lo digital

A lo largo de este capítulo he intentado dar una perspectiva sobre aspectos clave de los medios digitales y de su teoría que nos ayuda a atravesarlos y comprenderlos. En el próximo capítulo, conectaré estas características con

³⁹ Manovich, *Software Takes Command. Extending the Language of New Media*, International Texts in Critical Media Aesthetics 5 (Nueva York: Bloomsbury, 2013).



las tradiciones de la preservación que exploramos en el anterior. Al juntarlos, intentaré articular y desglosar lo más claro posible algunas de las suposiciones tácitas que dan vida al trabajo en la preservación digital. En cuanto a la teoría, en este capítulo, quisiera volver a una de las implicaciones fundamentales de sus estudios: los medios digitales son desordenados.

Los comerciales de Apple presentan una retórica visual de la computación como un mundo blanco y esterilizado en el que las cosas “simplemente funcionan”. Ésta es la fachada de la computación. La realidad es mucho más parecida al mundo imperfecto lleno de cables y alambres que vemos en *The Matrix*.⁴⁰ Esta representación es indicativa del modo en que las superficies (las pantallas de las computadoras) ocultan de forma intencional las complejidades de la naturaleza de las plataformas y de la información escondida dentro de ellas. Hacen que el mundo digital parezca un universo genial, regulado por la lógica y el orden. En la práctica, es un mundo de especificaciones incompletas y contradictorias, de archivos que no son válidos, pero que parecen funcionar en las aplicaciones que la mayoría de la gente usa y, muy en el fondo, toda esa información está compuesta de marcas físicas en medios tangibles.

Esos medios se gastan. Son quebradizos. Se pueden deformar y degradar al calentarse. Se cubren de pelusas cuando el ventilador trata de enfriarlos. El desorden de lo digital está incluso escrito ahí en el código fuente. Como señala Matthew Kirschenbaum: “Los programadores tienen un nombre especial para la forma en que el *software* tiende a acumularse como capas de historia sedimentada, de fósiles y reliquias de versiones pasadas y áreas en desarrollo sin salida: ‘cruft’, una palabra tan llena de textura como costra o polvo y a otras que se refieren físicamente a cáscaras y sobras”.⁴¹ En cada nivel, las plataformas que permiten, limitan y construyen nuestras interacciones con la información digital están hechas a mano por individuos, compañías, comités internacionales. Casi cualquier aspecto de esa información puede decirnos cosas sobre las personas que guardaron los archivos, las que crearon el *software* y las comunidades y sociedades a las que pertenecen. El mundo digital es desordenado. Es importante que en nuestros esfuerzos por recopilar, preservar y dar acceso a él, no lo ordenemos en formas que le quiten partes importantes.

40 Para una exploración crítica de las diferencias entre las visiones estéticas de los comerciales de Apple y *The Matrix*, ver Lisa Nakamura, *Digitizing Race. Visual Cultures of the Internet*, Electronic Mediations 23 (Mineápolis: University of Minnesota Press, 2008).

41 Kirschenbaum, “Software, It’s a Thing”, *Medium*, 25 de julio de 2014, <https://medium.com/@mkirschenbaum/software-its-a-thing-a550448d0ed3>.

Capítulo 3

Los desafíos y las oportunidades de la preservación digital

Con algo de conocimiento sobre los medios digitales y un poco de contexto sobre varias tradiciones de preservación, ahora es posible desglosar cuáles son los retos, oportunidades y suposiciones inherentes a la preservación digital. Una de las grandes lecciones en cuanto a las tradiciones de preservación es que siempre se trata de un diálogo entre las posibilidades materiales de nuestros medios y nuestra visión sobre las expresiones culturales perdurables. Nuestras formas culturales se remedian y reconfiguran a través de un diálogo con las posibilidades que ofrecen los medios y los formatos. Entonces, cuando se trata de la preservación digital, ¿dónde la ubicamos dentro de las tradiciones de preservación?

Como pasa con frecuencia, la respuesta es que es complicado. El enfoque de la tradición artefactual de las máquinas computacionales está fuera de la práctica convencional de la preservación digital. Las computadoras son sorprendentemente buenas para hacer copias de la información, tanto que cuando las hacen, algunos aspectos del artefacto terminan por ocurrir también en las capas de la plataforma de los objetos informativos. Intentaré describir de manera rápida lo que quiero decir con esto y, al mismo tiempo, aclararé cómo funciona este capítulo para explicarlo más a fondo.

Por diferentes razones que se expondrán adelante, el enfoque de la proximidad histórica de los artefactos ha sido abandonado de la corriente principal del trabajo de la preservación digital. El desafío de mantener los objetos digitales funcionando en sus medios originales es demasiado grande. Como consecuencia, la preservación digital trata con objetos informativos y no con el material subyacente en el que están codificados. Esto conformará la primera gran porción del capítulo.



Los medios digitales, afortunadamente, ofrecen grandes oportunidades. La informática se ha desarrollado alrededor de la creación y autentificación de copias. Por ello, nunca ha habido otro medio que esté tan preparado para el enfoque que tiene la preservación informativa. Así, dedicaré algo de tiempo en este capítulo para explicar más sobre este aspecto de la información digital.

Dicho esto, gracias a la naturaleza de plataformas anidadas de la información digital descrita en el último capítulo, existen todo tipo de características de los objetos digitales que terminan teniendo cualidades artefactuales. Esas capas de información son palimpsestos por naturaleza. Si bien la palabra “información” suena como un concepto demasiado directo y estéril, cuando profundizamos en todas las capas de las plataformas, encontramos una y otra vez información que parecía oculta. De esta manera, gran parte del razonamiento sobre preservación que es utilizado para trabajar con artefactos, en el enfoque de la proximidad histórica para tratar con objetos tangibles, es muy útil para abordar la preservación del contenido informativo en las cadenas de bits. Las posibilidades físicas y tangibles de los diferentes medios digitales continúan formando y estructurando el contenido digital mucho después de que se haya transferido y migrado a nuevos medios. Me doy cuenta de que esto es confuso. Se volverá mucho más claro a través de una serie de ejemplos del mundo real.

Entonces, existen bastantes cualidades artefactuales de los objetos informativos y, además, la preservación digital ha renunciado a la conservación de los medios físicos subyacentes. El resultado de estos dos puntos es que la preservación digital tiene una relación significativamente diferente con las ciencias de la preservación. Un conservador que trabaja con una pintura quiere comprender la química de todo. Para el conservador, la base del trabajo está en las ciencias del mundo natural y físico (química, ciencia de los materiales, física y biología). En contraste, las ciencias subyacentes para la preservación digital son las de lo artificial: las ciencias de la computación, la informática, el aprendizaje automático y la inteligencia artificial. Estamos comenzando a comprender las implicaciones de esto, pero eso significa que nuestras relaciones con las ciencias subyacentes deben cambiar a medida que avanzamos cada vez más para trabajar con la preservación digital.

La conservación de los medios digitales: inviable

Se ha derramado mucha tinta sobre por qué la preservación digital es tan difícil. Gran parte del impulso de este trabajo proviene del anhelo por las posibilidades de los medios del pasado. Gracias a la larga vida y a las buenas condiciones alrededor de los materiales, hoy tenemos acceso directo a pinturas rupestres de



hace 40 mil años y a pergaminos, esculturas y edificios de hace dos mil años. Incluso en la historia más reciente, uno de los éxitos de la preservación, la microfilmación, fue la idea de que, en buenas condiciones, la copia de registros en ese material podría durar casi un milenio.

Las estimaciones sobre el promedio de vida de diferentes tipos de medios digitales varían dramáticamente, pero una cosa está clara: éstos no se parecen en nada a lo que tenemos con los medios analógicos. La cinta magnética, los disquetes, los discos duros, el almacenamiento *flash*, los CD y los DVD son diferentes, en cierta medida, pero el punto central es que, en general, estos medios están bien durante una década o dos y después requieren de algunas técnicas bastante extremas para extraerles datos. Es verdad que ha habido una gran variedad de intentos para crear medios más longevos: el M-Disc es un CD de piedra que pretende estar bien por mil años; hay un esfuerzo por crear un almacenamiento de cinta basado en aluminio que mantenga los datos seguros incluso en un incendio. Sin embargo, estas tecnologías aún no han despegado realmente en el mercado. Los expertos son muy escépticos y dudan que incluso lleguen a despegar algún día porque el ciclo de actualización de las tecnologías es tan rápido que la mayor parte de los medios digitales nunca se acercan al final de su vida útil antes de que sean descartados.

Las interfaces, tanto las conexiones de *hardware* (piense en lo que se necesitaría para encontrar un lector para un disquete de ocho pulgadas o para un disco Zip) como los programas de *software* y los controladores que pueden abrir y leer datos, van y vienen mucho antes de que los medios hayan llegado al final de su vida útil. Ésta no es una característica propia de la informática. Tiene más que ver con la obsolescencia programada de los medios y la gigantesca industria informática que vierte recursos considerables para mantener el ritmo de las expectativas de los dispositivos informáticos cada vez mejores, más rápidos, con más almacenamiento y cada vez más pequeños, que con su naturaleza. Ahora, si bien hemos visto un crecimiento exponencial en la cantidad de información digital que podemos almacenar en un espacio físico determinado durante el último medio siglo, no existe una razón inherente para que esto continúe al mismo ritmo. Cada avance sucesivo en la tecnología del almacenamiento digital ha implicado nuevas técnicas sobre cómo aumentar la capacidad de los medios de almacenamiento y mantener los tiempos de lectura altos, pero existen límites físicos muy reales para algunos medios a los que estamos por llegar. En cualquier caso, los resultados para la preservación de los medios digitales siguen siendo los mismos. Incluso si se tienen medios antiguos, se necesita mucho esfuerzo para poder acceder a la información. Ese trabajo



se vuelve cada vez más difícil a medida que pasan los años y los medios subyacentes se siguen degradando.

A lo largo de la historia, los conservadores desarrollaron y crearon técnicas para reparar, manipular o hacer que los medios vuelvan a servir. De esta forma, si se tiene un trabajo en papel que no es libre de ácido, se puede desacidificar para extender su vida útil. Del mismo modo, en la restauración arquitectónica, un conservador podría incluso colaborar en un proyecto para restaurar parte de una estructura sólo con base en una comprensión profunda de cómo se construyó. Esto es inimaginable en el caso de los medios digitales. Un disco duro, por ejemplo, es un dispositivo de una complejidad increíble, hecho de una variedad de partes que serían casi imposibles de conseguir, y las especificaciones para tales dispositivos, con toda probabilidad, serán igual de difíciles de adquirir. Por lo tanto, casi no hay esperanza de que exista un futuro en el que un conservador de discos duros pueda ocuparse de reparar un medio de 50 años para conectarlo y que el usuario tenga acceso a él.

Las nociones tradicionales de la ciencia de la conservación, fuera de algunos casos específicos, son inútiles para la preservación a largo plazo de los objetos digitales. Con esto señalado, es fácil caer en exageraciones. La mayor parte de las ideas sobre la longevidad de los medios digitales son propiamente teóricas; se basan en pruebas para simular su uso y no en el desgaste real de los medios antiguos. A medida que los archivistas han comenzado a utilizar cada vez más técnicas forenses, muchos están descubriendo que pueden extraer datos de medios que tienen tres o cuatro décadas de antigüedad. Del mismo modo, se están desarrollando herramientas y técnicas para leer y tratar con medios dañados o degradados.

Mientras que mantener un *hardware* de computación que va quedando en desuso está fuera del alcance de la práctica de la preservación, ha surgido un área de estudio vibrante (e invaluable) para trabajar con estos dispositivos en específico. Los laboratorios de humanidades digitales y nuevos medios, como Trope Tank en el Massachusetts Institute of Technology (MIT), el Media Archeology Lab de la University of Colorado Boulder, la Media Archaeological Fundus (MAF) en la Humboldt University y el Maryland Institute for Technology in the Humanities (MITH) en la Maryland University han establecido colecciones de tecnología informática histórica, que han permitido la exploración y el análisis extensos de estas tecnologías.⁴² Es poco probable que estas máquinas

⁴² Para una exploración más profunda de las conexiones entre los diferentes laboratorios de medios arqueológicos, ver Lori Emerson, "Archives, Materiality, and the 'Agency of the Ma-



funcionen dentro de cien años, pero todavía existe un espacio de tiempo para que los investigadores trabajen con ellas y nos ayuden a entenderlas mejor. La labor que se lleva a cabo en estos laboratorios está generando un entendimiento hacia estas tecnologías que será clave para trabajar con las máquinas. La investigación será instrumental en nuestra habilidad para comprender el contenido que se pasará a nuevos sistemas y que también contribuirá a saber cómo superar el contenido. Este trabajo es relevante y significativo, pero representa un área de trabajo y práctica diferente. Hoy en día está aceptado que, si es posible recuperar datos de esos medios, el primer paso en la preservación es copiar la información de ese medio y moverla a un sistema de almacenamiento digital contemporáneo. El futuro de la preservación digital no está en el mantenimiento de los antiguos sistemas de informática y *hardware*. Las prácticas de la preservación digital están sobre todo encaminadas a hacer copias y a mover esas copias hacia nuevos sistemas.

La buena noticia es que hay una cosa para la que las tecnologías informáticas están diseñadas y es precisamente para copiar.

La informática se trata de copiar

La preservación de los objetos informativos (aquellos que se suscriben al enfoque de identidad con las “mismas letras”) siempre ha requerido cierta fe en la longevidad de los medios subyacentes. Las palabras están escritas en el pergamino, es decir, habitan en sus medios. Las palabras esperan ser copiadas y transferidas en una fecha futura. Copiar siempre ha sido un poco complicado, también. ¿Qué tan seguro está usted de que esta copia es idéntica a la última en todas las formas significativas? De hecho, gran parte del trabajo en el estudio de manuscritos se centra en cómo las copias divergen con el tiempo. Siempre ha habido una cantidad de errores humanos que se arrastran en ese proceso. Los medios informáticos son, en principio, diferentes a este respecto.

Reproducir un archivo, en términos generales, implica copiarlo, es decir, el archivo existe en un medio de almacenamiento, pero es leído y transferido de nuevo para poder ser visto. En realidad, uno no ve la copia en el disco duro, sino una copia transferida a otra parte de la computadora. Esto también es cierto para cualquier cosa que se vea en la red. Para que cualquier cosa

chine’: An Interview with Wolfgang Ernst”, *The Signal* (blog), 8 de febrero de 2013, blogs.loc.gov/thesignal/2013/02/archives-materiality-and-agency-of-the-machine-an-interview-with-wolfgang-ernst/.



aparezca en la pantalla, primero debe ser copiada de su fuente a la computadora. Crear, transferir y verificar copias es, por lo tanto, una parte central de la naturaleza de la informática.

Las computadoras están diseñadas, en gran medida, para tratar sus medios de almacenamiento más como un armario (donde se intenta organizar y almacenar las cosas de manera tan eficiente como sea posible) y menos como un medio en el que están escribiendo una historia (donde se vuelve al lugar en que se quedó para comenzar de nuevo). Si usted alguna vez ha desfragmentado su disco duro, sabrá que es muy parecido a reorganizar un armario. A medida que su computadora ha escondido archivos aquí y allá a través de sectores en un disco duro, vuelve y trata de ordenarlos, para optimizar el lugar donde se almacena cada uno. Sí, en el nivel más básico, la computadora escribe, sobrescribe y reescribe sobre el medio, pero el sistema lo trata más como un contenedor y menos como un medio. Acercarse a un disco duro como un contenedor es hacerlo como lo haría una computadora.

Conforme la información se mueve de un punto a otro, se va transfiriendo entre computadoras, o dentro de los mismos componentes internos de una, y su integridad es verificada y vuelta a verificar de manera más o menos constante. La mayoría de las redes digitales y los dispositivos de almacenamiento calculan las “comprobaciones de redundancia cíclicas” (CRC) para confirmar que la información se transfiera y copie correctamente. Estas comprobaciones calculan números que actúan como una especie de huella digital para el archivo, de modo que se pueda volver a calcular una vez que se transmite para identificar cuando el archivo no se haya copiado con exactitud.

En esta área, los ingenieros en informática han desarrollado una amplia gama de estrategias y técnicas para hacer que los datos sean más perdurables frente a la desconfianza en la longevidad de los medios físicos y la necesidad de copiar y transferir información constantemente. Lo que originalmente significaba “variedad redundante de discos baratos”, ahora conocido como RAID, se diseñó en 1987. Su construcción se basó en estrategias pensadas más de una década antes. Esta tecnología ofrece una forma de virtualizar un sistema de almacenamiento para mantener múltiples copias en los medios subyacentes; así, si una unidad falla, el contenido puede persistir en las otras unidades. Del mismo modo, la corrección de errores hacia adelante, técnica de codificación de información redundante, es decir, de múltiples copias de la información en el objeto mismo, se remonta a la década de 1940. En este caso, la información redundante se codifica con el objeto transferido, de modo que, si ésta se degrada o se pierde durante la transmisión, es posible reconstruirla para que sea



idéntica a la que se envió. Otro ejemplo es el sistema de archivos ZFS, cuyo nombre original era “sistema de archivos Zettabyte”, y otros similares que computan las verificaciones al nivel de bloque del dispositivo de almacenamiento y, sin que el usuario lo sepa, verifican y reparan los datos. Estas técnicas demuestran cómo los sistemas informáticos consideran el almacenamiento.

En la preservación digital, este tipo de revisión se llama “verificación de integridad”. Al comparar la huella digital de un objeto con un archivo en el futuro, podemos ofrecer evidencia indudable de que lo que tenemos es lo mismo que teníamos antes. Hay muchas formas de generar fijeza de la información: desde asegurarse de que el archivo tenga el mismo tamaño hasta calcular las CRC o usar funciones criptográficas *hash* más sofisticadas, que fueron desarrolladas para poder garantizar la autenticidad de la información digital, en caso de que alguien intente manipularla. Todos estos métodos ofrecen evidencia de que la información de un archivo digital (la secuencia completa de ceros y unos) es exactamente igual a la que era antes. Los que trabajamos en la preservación digital obtenemos toda esta tecnología y capacidades de manera gratuita.

El cuidado y la preservación de los objetos informativos han existido durante milenios. Las culturas han copiado durante mucho tiempo poemas, historias y otras obras. Esas copias siempre han estado un poco en desacuerdo con los medios. Sin embargo, ahora ya no es así. Las ciencias de la computación y la informática han necesitado diseñar e implementar técnicas integradas en el núcleo de la tecnología que, en esencia, se centran en cómo hacer y verificar copias perfectas. De esta manera, podemos adoptar la tradición informativa para la preservación y afirmar con confianza que nunca antes ha habido un medio tan bien integrado con la necesidad de preservar la información al hacer muchas copias y verificar que sean exactas.

Hasta este punto, podría parecer sencillo darle la espalda a la tradición de la preservación artefactual o de la conservación. Sin embargo, es un poco más complicado que eso. Si bien los preservadores digitales, en principio, han renunciado a los medios subyacentes, es curioso señalar que muchas de las posibilidades y tipos de preguntas con las que hemos podido abordar el beneficio de la materialidad de los objetos han abierto un camino hasta las capas superiores de estas plataformas. Así, hay una variedad de cualidades extrañamente artefactuales escondidas en los rincones y grietas de nuestros objetos informativos basados en plataformas.



Las cualidades artefactuales de los objetos informativos

Cuando hablábamos del libro de *Frankenstein*, fue hasta cierto punto transparente para todos entender qué es lo que cuenta como objeto informativo. Se puede ver cada letra, palabra y signo de puntuación. Cuando se copia todo eso en otra cosa, se obtiene la información del texto. Por el contrario, la naturaleza de las plataformas de la información provoca que exista información codificada en varios lugares diferentes que podría ser relevante, pero que no siempre es visible. Lo que nos trae de nueva cuenta a la idea del esencialismo de pantalla. Volviendo a uno de los ejemplos discutidos anteriormente, el *Palimpsesto de Arquímedes* termina representando un patrón que se repite una y otra vez en el mundo completamente informativo de los objetos digitales.

Recordemos que el *Palimpsesto de Arquímedes* es una copia griega bizantina de obras desconocidas del autor, que data del siglo X y que fue sobrescrita con un texto religioso en el siglo XIII. Mediante el uso de imágenes multispectrales y herramientas computacionales, fue posible leer el texto que se encontraba debajo. El objeto material retuvo rastros del texto de Arquímedes que no eran visibles para el ojo humano, pero que, de hecho, todavía estaban allí, en espera de ser descubiertos mediante la tecnología de escaneo de imágenes. Este patrón ahora se reproduce repetidamente, pero a través de los objetos informativos, lo que representa el punto central del esencialismo de pantalla descrito en el capítulo anterior.

A continuación, describiré de manera breve una serie de contextos: los archivos de palabras de un dramaturgo, la historia de dos copias individuales del videojuego de Apple II y algunos aspectos de un videojuego más reciente. Cada una de estas historias demuestra diferentes maneras en las que los objetos digitales informativos pueden llegar a portar cualidades artefactuales y, en algunos casos, incluso folclóricas. Cada caso también vuelve a los aspectos clave de la naturaleza de las plataformas de los objetos digitales, ya que estas cualidades podrían ser significativas y quedar ocultas a la vista a través de las interacciones entre las diversas capas de los *software*, los formatos y los protocolos necesarios para representar e interactuar con dichos objetos.

Los archivos de Word de Jonathan Larson

El compositor y dramaturgo Jonathan Larson, conocido por su musical *Rent*, murió de manera trágica poco antes del estreno de la obra en 1996. Sus papeles, junto con cerca de 180 disquetes, fueron regalados a la Library of Congress Music Division. El trabajo con estos archivos, del curador Doug Reside, se ha



convertido en uno de los ejemplos más significativos del tipo de estudio digital que se vuelve posible a medida que más y más colecciones de manuscritos comienzan a proporcionar acceso a este tipo de información de origen digital.⁴³

Larson escribió *Rent* con Microsoft Word 5.1 en una Mac. Si usted ejecuta una copia antigua de Word en una Mac o en una versión emulada de una Mac, podría leer el borrador final de la obra. Sin embargo, si abre los mismos archivos con otro editor de texto, verá algo diferente por completo. Muchas de las palabras son totalmente distintas. En la Mac se ve el texto como el autor lo habría visto, el borrador completo. Sin embargo, en otro editor se muestra algo similar, pero que a todas luces es diferente. Varias de las letras de las canciones del musical aparecen distintas de como se verían en Word. La versión de Word coincide con la partitura final. ¿Cómo puede ser que distintos procesadores de texto presenten texto diferente?

Word 5.1 hizo algo que llamó un “guardado rápido”. Cuando un usuario guarda un documento, en lugar de sobrescribir el archivo, se agrega todo el texto al final de él. Debido a que el contenido está codificado utilizando especificaciones de texto comunes, usted puede abrir y leer en otro editor, pero no verá lo que Larson habría visto cuando estaba editando su guión. Se verá el texto de un borrador anterior y luego, en la parte inferior del archivo, el que se incluye para sobrescribir esos puntos en particular en el texto que reemplaza. La lógica para dar sentido a esos fragmentos agregados que se guardan rápido en el archivo y para volver a incorporarlos puede no estar en un editor de texto, pero cuando se abre en otro, es posible ver parte de la historia de la creación del archivo. Escondida en él, invisible a la forma en que se utilizó, se puede encontrar la historia de las ediciones y los cambios en el texto. Si entender la historia de los borradores de este trabajo es importante para usted, entonces ha tenido suerte. Una gran cantidad de información de la historia de este texto está disponible para que la vea.

De acuerdo a la forma en que funcionó esta versión particular de Word, los documentos creados con ella vienen con estos rastros históricos. Cuando se utiliza el *software* correcto o, como en este caso en particular, el incorrecto, esas huellas salen a la luz para mostrar el proceso a través del cual se creó el documento.

43 Mi discusión sobre los documentos de Larson presenta los resultados de Doug Reside, “‘No Day but Today’: A Look at Jonathan Larson’s Word Files”, *New York Public Library Blog*, 22 de abril de 2011, <https://www.nypl.org/blog/2011/04/22/no-day-today-look-jonathan-larsons-word-files>.



El misterio en Mystery_house.dsk

Mystery House, un juego de 1980 para Apple II, le permite al jugador explorar una mansión victoriana abandonada. El juego presenta a los usuarios con imágenes que muestran diferentes habitaciones y lugares de la casa, así como un texto que describe lo que sucede a medida que se avanza. Los detalles no son particularmente importantes para este ejemplo.

Después de descargar una copia de la imagen de disco del juego, es decir, una copia de un disquete original de 5 1/4 en el que se guardó, es posible iniciarlo en un emulador de Apple II y examinarlo. Usted también puede tomar una copia como esa y explorarla a través de un editor Hex, el cual, como expliqué antes, es un programa que permite leer los datos binarios reales presentados en un medio determinado. Hasta este punto, el disco sólo es un objeto virtual. El archivo de imagen del disco es una copia bit por bit de cómo estaba la información en algún disco real, pero no tenemos idea de dónde está ese disco físico original o si aún existe.⁴⁴

Cuando Matthew Kirshenbaum hizo justo eso, es decir, descargó una copia de la imagen de disco del juego desde la red y comenzó a mirar alrededor con un editor Hex, terminó encontrando algo divertido. Como recordatorio, el texto codificado dentro de un archivo es en sí mismo interpretable y representable desde el exterior. Entonces, al explorar sectores de un disco con un editor Hex, éste puede representar el texto tal y como está desplegado en el disco. Lo que Kirshenbaum encontró fue texto de dos juegos diferentes por completo: *Dung Beatles* de 1982 y *Blitzkrieg* de 1979. Como sucede con muchos medios de almacenamiento digital, la información en un disco no se borra cuando se elimina, sino que el espacio sólo se marca como vacío. En consecuencia, cuando se hurga en los sectores reales del disco o de la copia bit por bit de éstos en una imagen de disco, se pueden encontrar rastros de archivos que se sobrescribieron. El resultado es que algunas décadas más tarde, al explorar los sectores de la copia del disco, era posible apreciar lo que había antes de que *Mystery House* se guardara.

La imagen de disco del juego es completamente informativa; es una secuencia de bits que se ha copiado y compartido con muchos usuarios. Sin embargo, en el proceso de copiar el disco, en realidad, se copió más que el juego. Al explorar el contenido del espacio que parece estar sobrescrito en el disco, es posible descubrir los usos anteriores del objeto físico en el que el juego había

⁴⁴ Para un recorrido forense más detallado de este disco, ver Kirshenbaum, *Mechanisms*, 114-142.



sido codificado en su momento. Los aspectos artefactuales de ese objeto físico se han llevado al mundo informativo de la replicación.

“Aquí yacen el peperony y el queso”

Otro ejemplo de un videojuego sirve para reafirmar la forma en que los aspectos personales e individuales del uso y de las experiencias con los medios digitales pueden codificarse en ellos y transmitirse posteriormente, en algunos casos, incluso adquiriendo su propia vida cultural. *Oregon Trail* es un videojuego educativo lanzado en una variedad de representaciones y versiones que datan de 1971. Es un poco icónico y nostálgico para las personas que crecieron en Estados Unidos en los años 80 y 90.

La mayoría de los tropos asociados con la nostalgia del juego son producto de la intención de sus diseñadores. Por ejemplo, la frase frecuente “Has muerto de disentería” es algo que los programadores escribieron en el juego. Sin embargo, al menos uno de los tropos y memes asociados con él fue generado por un usuario, y su persistencia atestigua la historia generalizada de copias piratas del juego.

Aquellos que crecieron con *Oregon Trail* recordarán cómo, con el tiempo, el escenario del juego se llenó de lápidas conmemorativas de diferentes jugadores que participaron previamente. Esas lápidas son bastante populares; incluso existe un generador gratuito en la red en el que se pueden crear y compartir con amigos. Lo fascinante de las lápidas era que persistían a través de los juegos individuales de persona a persona. Lo que alguien escribía en una lápida para su personaje aparecía en los juegos de futuros usuarios de esa copia particular del juego. Es como un tipo de marcador sombrío que creó una capa de interacción fuera de las partidas individuales. Además, hizo que el juego funcionara como un mundo en curso. Las tumbas representaron el paso del tiempo en ese mundo. Cada juego que sucedía después del anterior iba apilando más lápidas con sus mensajes, a veces absurdos, por el camino.

Gracias a la piratería, una de esas lápidas adquirió una vida cultural propia. Cada mensaje de lápida se guardó en una parte del disco en el que se ejecutaba el juego. En lugar de comprar copias originales limpias, muchos jugadores usaron copias pirata. Cuando arrancaban una de esas copias, encontraban una lápida que decía: “Aquí yace andy; peperony y queso”. Como contexto de la broma, la marca de pizza Tombstone publicó un comercial, en los años 90, en que se le preguntaba a un alguacil a punto de ser ejecutado qué quería en su *tombstone*, es decir, su tumba, a lo que contestaba con los ingredientes que le gustaría en su pizza, en este caso, peperoni y queso.



Entonces, un jugador, presumiblemente llamado Andy, escribió una broma con faltas de ortografía sobre este comercial en su copia del juego y luego dejó que otros la copiaran. Esa copia se convirtió en la base de un montón de copias pirata y, en consecuencia, la lápida que el jugador escribió en su copia local se ha convertido en una conmemoración que cobró vida propia como un meme fuera de ese mundo.⁴⁵ La información guardada en esa copia particular se ha convertido en parte del significado cultural del juego.

Este ejemplo ilustra las formas en las que los objetos folclóricos que, en gran parte han estado ausentes de esta discusión, siguen formando parte de nuestro ecosistema de preservación digital. En este caso, una broma de uno de los usuarios entra en la historia clásica del juego. Una alteración individual en el juego surgió y se reprodujo de manera muy similar a las variaciones replicadas en las tradiciones orales y manuscritas.

La representación de los pueblos originarios en la colonización

El lanzamiento en 2008 de *Sid Meier's Civilization IV: Colonization* llegó con mucha controversia. En el juego, los usuarios representan el papel de un poder colonial que domina, saquea y conquista América. Si bien la superficie del juego, es decir, la manera en la que muestra, representa y modela el encuentro colonial está disponible para su análisis y crítica, la forma en la que está construido en sí mismo, también abre otra vía para abordarlo: el código fuente.

Civilization IV se diseñó para poder ser modificado. Gran parte del juego está escrito en scripts de Python que son legibles y modificables por sus usuarios. Si bien hay una parte central que ejecuta código compilado que no se puede leer como texto, muchas de las reglas y la lógica están hechas de manera que los jugadores puedan editarlas y cambiarlas. *Colonization* se lanzó comercialmente y es, en términos de funcionalidad, una modificación grande del juego original. Utiliza su motor central e incluye muchos archivos adicionales que pueden reescribir algunas de las reglas para cambiarlo.

La investigadora Rebecca Mir y yo pudimos explorar los archivos del juego para descubrir cómo establece las diferencias entre los pueblos originarios de América y los europeos.⁴⁶ Desde la perspectiva del código fuente, hay pueblos

⁴⁵ Este ejemplo proviene de un escrito anterior mío sobre las representaciones de monumentos conmemorativos en videojuegos, ver Trevor Owens, "Pixelated Commemorations: 4 In Game Monuments and Memorials", *Play the Past*, 18 de junio de 2014, <http://www.playthepast.org/?p=4811>.

⁴⁶ *Ibid.*



“normales” (unidades coloniales controladas por el jugador), pueblos nativos y pueblos europeos (ambos controlados por la computadora). Las personas normales vienen con una gama de habilidades y características. El código quita muchas de ellas a quienes están marcados como “isNative”, es decir, nativos. Entonces, en el código fuente, se puede rastrear con precisión qué habilidades de las potencias europeas se eliminan de los pueblos nativos. Entonces, más allá de comprender cuál es la experiencia del juego, es posible ver con exactitud cómo representa esa visión colonial debido a la manera en que se diseñó y se configuró. De hecho, hay un área completa de estudios de humanidades, llamada “estudios críticos de código”, que se está desarrollando alrededor de hacer este tipo de lectura detallada del código fuente, en especial para explorar su significado extrafuncional.

Lo relevante aquí es que, gracias a la planeación del juego, cada copia venía con partes abiertas de su código fuente subyacente para que cualquiera las pudiera explorar y leer. Incluso si no se jugaba, se podía aprender mucho sobre cómo funcionaba leyendo los scripts de Python y los archivos XML que contenía. Así, el juego-objeto tenía la capacidad para experimentarlo, ejecutarlo y, cuando se exploraba a través de su sistema de archivos, para deconstruir una parte de la forma en que operaba.

La copia de cadenas de bits como una acción de preservación

Tomados en conjunto, estos cuatro ejemplos ilustran cómo las interacciones entre todas estas capas en la naturaleza de las plataformas de los objetos digitales crean lugares donde la información potencialmente significativa podría ocultarse o presentarse para ser explorada y analizada. Donde era hasta cierto punto sencillo definir las cualidades informativas de un libro (asegurarse de que todas las palabras estén escritas de la misma manera) hay una cantidad casi interminable de información que podría ser relevante y que está ligada a las cualidades artefactuales de los objetos digitales. Como resultado, muchos de los que trabajan en la preservación digital han comenzado a considerar la posibilidad de obtener una copia bit a bit de toda la información en medios físicos, no como una copia, sino como una acción de preservación, estabilización y conservación de un objeto físico. Desde esta perspectiva, nos movemos cada vez más hacia un conjunto diferente de relaciones en cuanto al apoyo de las ciencias para la preservación.

El cómic de XKCD, “Old Files”, ofrece un ejemplo práctico de cómo esto forma parte de la experiencia de preservar nuestro pasado digital. Muestra al protagonista, una representación del autor, en la parte inferior de la estratigrafía



de sus archivos personales digitales. A medida que recorre las capas, comenzando por lo más actual en sus “Documentos”, luego avanzando hacia su “Escritorio anterior” y, más adelante, su “Disco zip de la secundaria”, encontramos burbujas que representan varios conjuntos de archivos, fotos de un baile, una nota de amor. Al final, parece que abrió un archivo de texto y redescubrió que de joven había escrito poesía.



Figura 2. “Old Files”, de Randall Munroe. Tomado de XKCD: *A Webcomic of Romance, Sarcasm, Math, and Language*, 2014. Licencia Creative Commons — Atribución-NoComercial 2.5.

Este cómic retrata la forma en que se crean los nombres de archivo para un uso rápido, como “Escritorio anterior”, y que acaban por convertirse en partes permanentes de la organización de la información y de cómo la práctica de copiar archivos viejos se parece mucho a raspar capas de sedimentos y leer nuestro propio pasado. Lo artefactual se codifica en todas las capas y se transporta a través del tiempo conforme los datos pasan de un medio a otro.



La representación, la sustentabilidad de los formatos de archivo y los límites de los objetos permeables

Cada uno de los casos descritos aquí, los diversos tipos de documentos de texto y los videojuegos, tienen como uno de sus temas centrales la manera de abrir, ver e interactuar con un archivo. Un procesador de texto muestra cierta información, mientras que otro no. Esto llama la atención sobre un aspecto fundamental de la naturaleza de las plataformas de los medios digitales que es único y desafiante para fines de preservación. Un libro impreso, una pintura, una estatua, son objetos que experimentamos a través de los sentidos. Sí, se necesita saber cómo leer el texto del libro para que tenga sentido, pero no se necesita saber qué esquema de codificación se usó, ni tener las fuentes correctas instaladas, ni abrirlo en la versión correcta de WordPerfect.

La representación auténtica del contenido digital, sin confundir la que se da en una situación específica con todas las que son posibles, se convierte en un tema central para la consideración de la preservación digital. Como demuestran los ejemplos en este capítulo y el anterior, la especificación de los límites de los objetos digitales es en sí mismo un ejercicio intelectual desafiante, muy distinto a lo que era necesario para los discretos y en gran medida independientes objetos físicos. Por lo general, tener cadenas de bits no es suficiente para nosotros; debemos comprender los problemas relacionados con las probabilidades de poder reproducir trabajos en un formato determinado en el futuro y entender qué usos podríamos dar a los objetos para los que posiblemente perdamos la capacidad de hacerlo. Éstos son puntos a los que volveré varias veces a lo largo del libro.

Las ciencias de la computación como la ciencia de los materiales digitales

Si tomamos la cadena de bits como la capa base, el objeto que debe ser preservado y luego estudiado y analizado por los usuarios, entonces hemos entrado en una relación muy diferente con las ciencias de la computación. Cuando el Royal Museum de Berlín abrió un laboratorio para conservar colecciones en 1888, contrataron a un químico para que lo administrara. La ciencia de la conservación ha involucrado un profundo compromiso con las ciencias naturales y biológicas.

La conservación de los objetos físicos siempre se puede mejorar a través de una comprensión más profunda de su naturaleza. Sin embargo, en el momento en que dejamos el mundo de los objetos físicos, es decir, de los medios,



y nos movemos hacia la capa de las cadenas de bits, nos apartamos del mundo natural y entramos en uno donde se encuentran las ciencias de lo artificial, las de la computación, la informática y una variedad cada vez mayor de campos y disciplinas que se están desarrollando y se basan completamente en la premisa del trabajo y la manipulación de la información digital.

Así, el campo de la preservación digital debe desarrollar y explorar aún más las posibles relaciones con los enfoques e ideas de las ciencias de la computación. Por ejemplo, las tecnologías de virtualización y emulación, es decir, los enfoques para replicar de forma virtual tanto el *hardware* como el *software*, continúan avanzando y muestran una promesa considerable para el trabajo en la preservación digital. Las técnicas de visión por computadora y las de procesamiento del lenguaje natural están desarrollándose y presentan un gran potencial para la organización y la descripción de los objetos digitales.

Regresando a un punto que toqué al final del último capítulo, la información digital es mucho más desordenada de lo que parece que debería ser. Es todo unos y ceros ¿verdad?, sin mencionar que hemos configurado una amplia gama de herramientas para que sea más fácil asegurarnos de copiarlos correctamente. El desafío es que, encima de la capa base de la cadena de bits, las capas anidadas de las plataformas, todos los formatos de archivo, los estándares de codificación, los sistemas operativos, las aplicaciones de representación, etc., todo necesita estar conectado para permitir varias formas de interactuar y dar sentido a esa información. Por lo tanto, necesitamos ser proactivos al hacer juicios sobre lo que importa de un conjunto de materiales, los riesgos que estamos dispuestos a correr y la manera en la que podemos hacer que sean accesibles hoy, pero con un ojo puesto en el futuro a largo plazo. La preservación digital no tiene esperanza de ser una ciencia exacta. En cambio, debe ser un oficio fundamentado en las teorías que funcionan tanto para la preservación como para la información digital. En el resto del libro, intentaré ofrecer una guía práctica, funcional e iterativa sobre cómo abordar este oficio.

Capítulo 4

El oficio de la preservación digital

La preservación digital es un oficio, no una ciencia. Al menos, ése es el argumento central de este libro. Los capítulos anteriores establecen las bases por las que es así. La preservación digital no debe ser entendida como una ruptura con un pasado análogo imaginario, sino como un diálogo profesional continuo, aunque a veces antagónico, con las nociones de preservación que existen desde el surgimiento mismo de nuestra civilización. Además, la historia de la computación debe ser considerada en principio como un fenómeno cultural y social, y no como uno tecnológico. La computación nos es vendida, a través de la retórica, como una serie de mecanismos lógicos, limpios y eficientes, casi perfectos. Sin embargo, la experiencia nos muestra que las computadoras están inmersas en un desorden similar al que rige en nuestro mundo análogo. Las unidades de disco fallan. Los bits se desordenan. Los archivos se corrompen. El *software* se llena de elementos disfuncionales. La preservación debe ser un oficio y no una ciencia, primero, porque su práctica necesita basarse en un continuo e irresoluto diálogo con las profesiones de la preservación y, segundo, porque debe ser sensible al desorden inherente y a la naturaleza accidentada a lo largo de la historia de la lógica informática.

Al definir la preservación digital como oficio, afirmo que no se trata de soluciones universales, sino de la generación de un enfoque adecuado para cada contexto de preservación específico. La naturaleza de la preservación digital no es técnica. El acceso a la información digital a largo plazo requiere de enfoques flexibles de planeación que puedan interactuar de manera constante con asuntos conceptuales. La preservación no puede ser llevada a cabo sólo por máquinas, sino que requiere del trabajo de artesanos o maestros de oficio que, de manera reflexiva, puedan acercarse a los problemas de la preservación



digital *in situ* y que desarrollen estrategias con base en los diferentes recursos, materiales y limitaciones conceptuales de cada situación.

A continuación, intentaré establecer un marco de referencia para el oficio de la preservación digital. La intención no es delinear el conjunto de pasos secuenciales de un proceso. Más bien, ofreceré puntos de entrada a una serie de asuntos que deben ser considerados dentro del trabajo para el aseguramiento del acceso a largo plazo de la información digital, ya que diferentes tipos de contenido desarrollados para diferentes usos y diferentes audiencias requieren diferentes enfoques. Además, esas diferencias están fundamentalmente atadas y constreñidas a los recursos con los que cuenta cada institución. El oficio de la preservación digital es, en definitiva, un diálogo continuo entre las posibilidades de medios, las intenciones de los autores y curadores y de los recursos y compromisos institucionales dedicados a la preservación.

Los enfoques expansivo, iterativo e interactivo

La preservación es tener acceso a algo en el futuro. Todo lo necesario para permitir esto es por fuerza parte de la preservación digital. Requiere que tengamos una visión amplia de la materia. La preservación digital precisa de la planeación, la descripción y los modos de acceso adecuados. Existe un orden lógico para resolver estas cuestiones que proporciona la estructura organizativa del resto del libro. En ese sentido, los siguientes cuatro capítulos abordan una amplia gama de temas relacionados con garantizar el acceso en el futuro. Cada uno se enfoca en un área de oficio: 1) el propósito de la preservación y el desarrollo de colecciones, 2) la gestión de copias y formatos, 3) la organización y descripción de contenido, y 4) los accesos multimodales.

En cuanto al lenguaje de este capítulo, he preferido usar la terminología archivística (valoración, organización y descripción, etc.) por encima de la bibliográfica o curatorial. Varios aspectos del trabajo de archivo resuenan con puntos comunes a muchas formas del contenido digital. Los archivistas han lidiado con relaciones jerárquicas complejas y con la suma de registros a escalas masivas durante mucho tiempo. Como resultado, el marco mental del archivero, a menudo, es el más adecuado de las profesiones relacionadas al patrimonio cultural (en bibliotecas, archivos, museos, conservación, folclor, historia, arqueología, literatura, gestión de registros y antropología, entre otras) para trabajar con la preservación digital.⁴⁷

⁴⁷ Jackie Dooley, *The Archival Advantage. Integrating Archival Expertise into Management of Born-Digital Library Materials* (Dublín, Ohio: OCLC Research, 2015), <http://www.oclc.org/content/dam/research/publications/2015/oclcresearch-archival-advantage-2015.pdf>.



Aunque utilice esta terminología, es importante recordar que las prácticas de formación de colecciones en bibliotecas, archivos y museos no son estáticas. A veces pareciera que simplemente necesitamos importar al entorno digital la manera en que se preserva y da acceso a las colecciones análogas de esas instituciones. En general, éste es un buen punto de partida. Los valores y la sabiduría de las profesiones relacionadas con estas prácticas no deben ser desechadas sólo porque algo que parece nuevo ha surgido. Sin embargo, esto puede sobre simplificar los procesos dinámicos y evolutivos del trabajo con colecciones análogas al verlos como algo estático contra el que comparar y contrastar, pero las prácticas de la preservación y el acceso al material análogo son todo menos eso.

De hecho, la preservación y el acceso al material análogo no es un problema resuelto. Se han producido ya varios cambios en las distintas tradiciones de preservación y siguen ocurriendo en este momento. Es fundamental asegurarnos de que la preservación digital entre en diálogo con lo más actual de todas las áreas de nuestros campos y no se base en ideas anticuadas sobre colecciones. Por ejemplo, el enfoque “más producto, menos proceso” (MPLP) tiene más de una década, pero no se ha introducido del todo en la práctica de archivo.⁴⁸ Del mismo modo, el desarrollo y el crecimiento de los archivos comunitarios están recibiendo cada vez más atención como contexto diferente para comprender la práctica de archivo.⁴⁹ Si usted no está familiarizado con estos cambios, no se preocupe, discutiré aspectos tanto de MPLP como de archivos comunitarios en los espacios que sea pertinente en los siguientes capítulos. La cuestión clave aquí es notar que las prácticas del trabajo con colecciones están en constante evolución. La preservación no será ni se convertirá en un problema resuelto. Está ligada a una variedad de ideologías y valores que ha evolucionado y continuará haciéndolo como pasa con cualquier oficio profesional. Las cosas no han sido preservadas, sino que *están siendo* preservadas. Siempre es un proceso activo. Resolvemos problemas particulares de preservación en contextos individuales y nos aseguramos de su entrega a las próximas generaciones de artesanos de la preservación.

48 Mark Greene y Dennis Meissner, “More Product, Less Process: Revamping Traditional Archival Processing”, *American Archivist* 68, núm. 2 (2005): 208-263.

49 Para más información sobre archivos comunitarios, ver Michelle Caswell, “Assessing the Use of Community Archives US IMLS RE-31-16-0117-16”, Institute of Museum and Library Services, acceso 28 de abril de 2021, https://www.imls.gov/sites/default/files/project-proposals/re-31-16-0117-16_proposal_documents.pdf.



Al separar cada una de las áreas del oficio y ordenarlas en los siguientes cuatro capítulos, me doy cuenta de que esto podría parecer un proceso secuencial directo. Hay una lógica en esta secuencia. En términos generales, tiene sentido pensar en ordenar las cosas desde su intención hasta su acceso. Sin embargo, quiero enfatizar que, en la práctica, éste sigue siendo un proceso bastante no-lineal, iterativo e interactivo. En cualquier situación, los distintos problemas y posibilidades de los objetos y medios se harán evidentes en un área del oficio que tendrá efectos significativos sobre las demás. Así es como, a medida que se comienza a trabajar en una de esas áreas y se avanza hacia las demás, a menudo se necesitará cambiar de enfoque, ya que habrá surgido un nuevo desarrollo en alguna otra. El esquema que propongo aquí, y en los siguientes capítulos, no es una lista de verificación. Piénsela más como una guía sobre las cuestiones a considerar durante el proceso de creación y ejecución de su estrategia para la preservación digital.

El siguiente diagrama muestra cómo veo la interacción de estas cuatro áreas del oficio. El propósito de la preservación y el desarrollo de las colecciones funcionan como la base del trabajo de la preservación digital. Todo lo que uno hace para permitir el acceso a largo plazo a la información digital debe basarse en una intención establecida que defina directamente lo que se selecciona y lo que se retiene.

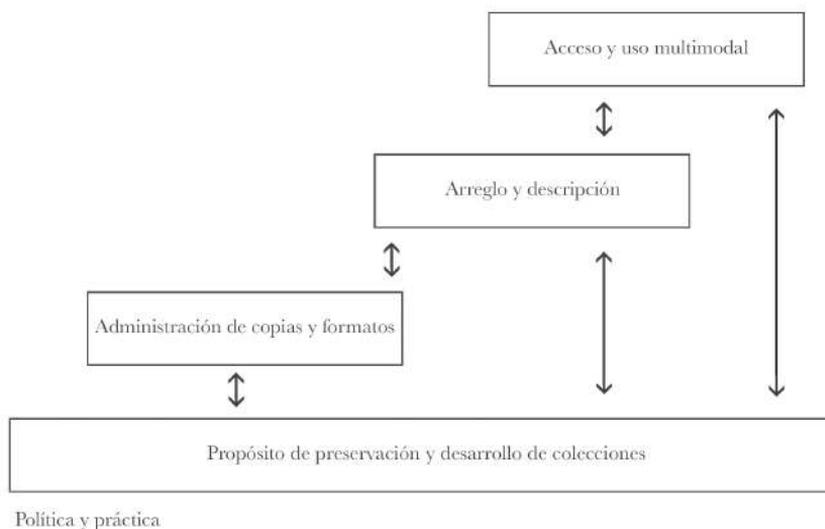


Figura 3. Áreas del oficio de la preservación digital y sus relaciones.

Elaborado por el autor. La traducción al español corresponde a la presente edición.



Después vienen las áreas centrales del trabajo de la preservación digital, la gestión de copias y formatos. El problema principal es cómo establecer sistemas y procedimientos que garanticen la preservación de bits y la planificación de una futura representación y usabilidad del contenido digital. Lo que sigue es organizar y describir. Sólo se tiene acceso al contenido cuando está organizado de manera legible para los usuarios actuales y futuros, y cuando se ha descrito de tal manera que pueda ser descubierto por esos usuarios.

Por último, he incluido los usos y accesos multimodales. Una de las posibilidades clave de los medios digitales es que el acceso es posible a través de una gama mucho más amplia de modos que en el caso de los medios analógicos. El acceso ya no necesita (ni debería necesitar) estar limitado a ver una sola página de una sola cosa en un solo lugar y momento. Estos nuevos modos de acceso traen consigo su propio conjunto de derechos, permisos y cuestiones éticas que también deben considerarse.

Todas estas áreas de trabajo funcionan como parte de una secuencia lógica. Sin embargo, he incluido flechas bidireccionales para subrayar que los problemas planteados en un área en particular pueden tener ramificaciones significativas para repensar los procesos en otra. Por ejemplo, si todo el contenido que preserva una institución no tiene restricciones de derechos o de privacidad en cuanto al acceso, esa organización podría diseñar su infraestructura de preservación central de manera mucho más abierta y con un acceso más amplio. Del mismo modo, si una entidad decidiera que todo lo que se necesita para acceder a una colección de un millón de imágenes es la capacidad de descargarlo en masa, tal como se recibe, esa decisión implicaría muchos menos recursos, que una colección para la que se haya determinado que es necesario desarrollar una descripción individual para cada imagen. En el primer caso podría tratarse, para fines de preservación, como un solo elemento con un único conjunto de valores de fijación, en vez de hacerlo como un millón de elementos individuales. A lo largo de los próximos capítulos destacaré los puntos en los que estas diferentes áreas interactúan o se conectan entre sí.

Bajo esa misma línea, es importante pensar, al menos en un nivel básico, en todas estas áreas en cada etapa del desarrollo de una estrategia o plan de preservación. Desde el comienzo de cualquier iniciativa de recopilación, es importante que piense en quién usará el material y cómo se lo va a hacer llegar. Dicho esto, no es necesario que usted tenga todo planeado por adelantado. Debe estar abierto a una gama de posibilidades que pueden surgir más adelante, a medida que las tecnologías con las que trabajamos continúan desarrollándose y evolucionando. Un buen trabajo de preservación digital se da al



enfrentarse a un problema determinado en todas estas áreas. Es deseable sacar a la luz los problemas que puedan surgir en cada una lo antes posible, y pensar en ellos de manera conjunta le ayuda a hacerlo. Todas estas áreas de actividad se establecen dentro de un marco más amplio. Cada una es habilitada y promulgada a través de la política y las prácticas de las instituciones que aseguran que se realice la preservación.

La política, la práctica y las instituciones

La política es lo que se dice que se va a hacer. La práctica es lo que realmente se hace. Ya sea que usted se encuentre en una biblioteca nacional con un personal de miles o en una casa-museo con sólo una persona, o se trate de un individuo que trabaja para mantener su información digital en orden, es importante que cuente con políticas y prácticas.

Se ha invertido una cantidad considerable de tiempo y energía en establecer políticas de preservación digital en algunas instituciones, así como en su revisión y su estudio.⁵⁰ Éste es un trabajo de vital importancia. Es esencial que una organización aclare lo que pretende hacer. Al interior, el desarrollo y el examen de las políticas sirven para aclarar qué roles y responsabilidades existen y cómo se distribuyen en las diferentes partes de la organización. En el exterior, la política de formalización ayuda a establecer expectativas sobre qué garantías ofrece la institución a sus partes interesadas. De cualquier forma, la política a menudo parece un poco más elegante de lo que en verdad es. En términos funcionales, tan sólo es una declaración de alto nivel sobre lo que se espera y lo que se lleva a cabo y, como tal, debe reflejar las realidades de la práctica.

La preservación sólo es posible a través de la alineación del trabajo de las personas, sus herramientas, sus flujos de trabajo y los recursos que se les da para apoyar esas actividades. Como tal, la política debe ser lo que alinee todas esas entradas y sean legibles para todos los involucrados. No he incluido un capítulo en el libro que se centre en la política o en la práctica, ya que todo lo que sigue en los próximos cuatro sirve como contribución que puede usarse de manera continua para refinar y definir las políticas y prácticas de una

50 Para obtener una descripción general de los temas en muchas de estas políticas, ver Madeline Sheldon, "Analysis of Current Digital Preservation Policies: Archives, Libraries, and Museums", *The Signal* (blog), 13 de agosto de 2013, <https://blogs.loc.gov/thesignal/2013/08/analysis-of-current-digital-preservation-policies-archives-libraries-and-museums/>.



organización. Con esto en mente, toda organización que se tome en serio la preservación digital debe establecer un proceso regular, al menos anual, mediante el cual se revisen las declaraciones explícitas sobre su política de preservación digital en función de los desarrollos continuos en cada una de las áreas del oficio.

La discusión sobre políticas de preservación me ha llevado a hablar mucho acerca de instituciones y organizaciones. Cuando digo “nosotros”, casi siempre me refiero al lector, es decir, alguien como yo, una persona lo suficientemente interesada en practicar la preservación digital como para leer este libro. Sin embargo, la preservación digital no es algo que puede hacerse a solas. Me doy cuenta de que algunos lectores pueden estar trabajando por su cuenta, fuera de cualquier organización o institución. No hay que desanimarse, más adelante comparto información que le será muy útil. Al menos durante un tiempo, cada uno de nosotros puede actuar como su propia institución o establecer una en nuestras comunidades. Éstas pueden interpretarse de manera tan amplia como para que incluyan familias, grupos cívicos, colectivos y cualquier otro tipo de asociaciones más informales. Los problemas descritos a continuación serán relevantes tanto para una organización de un solo individuo como para una que emplea a miles. No obstante, sólo aquellas instituciones que duren más tiempo que un individuo pueden consolidar el acceso a largo plazo.

La preservación se da gracias a las instituciones. Incluso si sólo se trata de organizar y administrar las fotos y los documentos familiares, se requerirá de algo más grande que usted para que persistan más allá de su tiempo de vida. Bien podría ser que se conserven a través de las instituciones familiares, entregadas y hechas legibles para la próxima generación, o a través de una sociedad histórica local. En cualquier caso, una persona sola no puede hacer la preservación digital. En el orden universal, nuestras vidas son fugaces. Dada la naturaleza de los medios digitales, es poco probable que, sin acciones e intervenciones, nuestros registros digitales duren mucho y sean inteligibles y accesibles sin el apoyo de una institución. Con ese fin, incluso para un individuo que trabaja solo, los capítulos que siguen son útiles para planear la entrega a alguna institución que será necesaria para garantizar el acceso al contenido en el futuro.

La relación con otros modelos y marcos

Existen diferentes modelos y marcos para la práctica y la planificación de la preservación digital. No es mi intención agregar más a esa lista. En particular, el Open Archival System Reference Model (OAIS), la Trustworthy Repositories Audit & Certification (TRAC) y, más recientemente, los Levels of Digital



Preservation (LODP) de la National Digital Stewardship Alliance (NDSA), ofrecen marcos para llevar a cabo la preservación digital.⁵¹ Éstos son sólo los más populares y los más utilizados en este momento, pero hay muchos más.⁵² Mi objetivo no es competir ni ofrecer otro adicional. Más bien, busco proporcionar un marco complementario para practicar de manera reflexiva la preservación digital como oficio. Los temas tratados en los siguientes capítulos deben ser considerados y explorados al trabajar con cualquiera de los modelos dados. Parte de la idea de la preservación digital vista como oficio es que no existe un sistema único para llevarla a cabo. Es un campo que requiere un refinamiento continuo. Los capítulos que siguen tienen el propósito de ayudar a las personas a enriquecer su oficio y a aplicarlo en apoyo a las misiones de las instituciones dedicadas al patrimonio cultural.

Mi enfoque está estrechamente alineado con los LODP, lo cual tiene sentido, ya que fui una de las personas que ayudó a desarrollarlos. La suposición central de los LODP es que la preservación digital se trata de hacer el mejor uso posible de los recursos, siempre limitados, para mitigar los riesgos más apremiantes de la pérdida de información digital a largo plazo. Vale la pena señalar que éstos cubren la mayor parte de los problemas que se plantearían en una auditoría TRAC. Entonces, la buena noticia es que existe un entendimiento considerable entre estos diferentes modelos y marcos.⁵³ De cualquier forma, debo aclarar que mi enfoque es un poco más amplio que los LODP, pues éstos

51 Para consultar una buena guía sobre OAIS, ver Brian F. Lavoie, *The Open Archival Information System (OAIS) Reference Model: Introductory Guide. DPC Technology Watch Report 14-02 October 2014* (Gran Bretaña: DPC, 2014), <http://www.dpconline.org/docman/technology-watch-reports/1359-dpctw14-02/file>; para obtener contexto sobre el uso de los estándares de auditoría TRAC, ver Bernadette Houghton, “Trustworthiness: Self-Assessment of an Institutional Repository against ISO 16363-2012”, *D-Lib Magazine* 21, núms. 3-4 (2015), DOI: 10.1045/march 2015-houghton; para más información sobre los niveles NDSA, ver Megan Phillips *et al.*, “The NDSA Levels of Digital Preservation: An Explanation and Uses” (presentación, IS&T Archiving Conference, 2-5 de abril de 2013), http://www.digitalpreservation.gov/documents/NDSA_Levels_Archiving_2013.pdf.

52 Para una muestra de seis modelos adicionales, ver Jefferson Bailey, “I Review 6 Digital Preservation Models So You Don’t Have To”, *Jefferson Bailey* (blog), 26 de abril de 2014, <http://www.jeffersonbailey.com/i-review-6-digital-preservation-models-so-you-dont-have-to/>; para más ejemplos, e incluso más diagramas, ver Bill LeFurgy, “Life Cycle Models for Digital Stewardship”, *The Signal* (blog), 21 de febrero de 2012, <http://blogs.loc.gov/thesignal/2012/02/life-cycle-models-for-digital-stewardship/>.

53 Amy Rudersdorf, “Proof Is in the Standards Pudding”, *AVP Blog*, 21 de junio de 2016, <https://www.weareavp.com/the-proof-is-in-the-standards-pudding/>.



se centran sólo en los aspectos o requisitos técnicos necesarios para la preservación digital. Los capítulos siguientes tratan sobre un conjunto más amplio de cuestiones requeridas para reflexionar sobre las políticas y prácticas de la preservación. No pasaré mucho tiempo discutiendo o trabajando estos marcos, pero los puntos que planteo en el resto del libro son útiles para trabajar dentro de cualquiera de ellos.

La concepción de la preservación digital como oficio es, en parte, una reacción a su propia “sobrediagramación”. Todos sus modelos están equivocados, pero algunos son más útiles que otros. Como tal, mi sugerencia es no apearse demasiado a ninguno de ellos, sino concentrarse en los temas centrales que están en juego en este trabajo. En mi experiencia, esto sucede en los casos en que el modelo OAIS se usa como la única lente a través de la cual, algunos en la preservación digital, pueden pensar el trabajo. Para algunos, las cajas en este modelo parecen ser rebasadas por las realidades prácticas que, en teoría, debería ayudar a sortear. Esto es importante en particular porque el propio OAIS establece una serie de relaciones de poder problemáticas entre creadores de contenido, usuarios e instituciones.⁵⁴ Los marcos son útiles como mapas, pero es fundamental recordar que un mapa no es el territorio. En ese sentido, los marcos sólo son útiles como herramientas en la medida en que ayudan a hacer el trabajo. Por lo tanto, no hay que tomar ninguno de ellos como un mandamiento que siempre debemos seguir, ni hay que detenerse demasiado en cómo se conceptualizan y se enmarcan los problemas de la preservación digital.

Los siguientes cuatro capítulos están destinados a proporcionar orientación al trazar ese terreno. Después de leerlos, usted estará bien posicionado para pensar qué es lo que importa sobre las cosas que desea preservar. Entonces, puede asegurarse de que su enfoque de preservación digital respeta lo que importa sobre ese contenido. A partir de ahí, puede estar seguro de que lo ha organizado y descrito de tal manera que pueda ser detectable e inteligible para sus usuarios ahora y en el futuro. Ofrecer a los usuarios acceso multimodal les permite dar retroalimentación, lo que a su vez lo ayuda a usted a tener un mejor entendimiento de la manera en que alguien puede hacer acceder y hacer uso de su contenido. Cuando usted realiza estas tareas, está llevando a cabo la preservación digital.

⁵⁴ Rhiannon Bettavia, “Encoding Power: The Scripting of Archival Structures in Digital Spaces using the Open Archival Information System (OAIS) Reference Model” (tesis de doctorado, University of Illinois at Urbana-Champaign, 2016).

Capítulo 5

El propósito de la preservación y el desarrollo de colecciones

Durante más de una década, millones de personas alrededor del mundo se han registrado para jugar en línea el videojuego de multijugadores masivo *World of Warcraft*. Cada vez que esos usuarios cargan una copia del juego en su computadora, éste se conecta a uno de sus múltiples servidores, en los que interactúan con miles de jugadores más. Juntos pueden completar misiones, combatir monstruos y enfrentarse en duelos.

¿Si usted quisiera preservar *World of Warcraft*, qué salvaría? Las copias individuales del videojuego sólo pueden ejecutarse cuando se conectan al entorno del servidor. Incluso si usted contara con un servidor configurado e iniciara sesión en el futuro, sería un escenario vacío sin sus jugadores. Gran parte de la experiencia de este juego está ligada a la cultura e interacción que surge entre sus usuarios, de tal modo que, aunque un grupo de personas ingresara dentro de cien años, nada de ese contexto social y cultural estaría presente.

Entonces, ¿qué es *World of Warcraft*? No existe sólo una respuesta. Hay muchos *World of Warcraft*. Es un paquete de *software* que se vende en las tiendas. Es el mundo que surge de la combinación de un *software* y un entorno de servidor. Es la comunidad y la cultura de los jugadores, documentadas en videos generados por los usuarios y por los sitios web auxiliares del gremio. Existe también como un objeto de diseño, ubicado en los registros de los recursos nativos digitales de Blizzard Entertainment, la compañía que lo creó y que lo ejecuta. Si estuviera interesado en la historia de la compañía, esto bien podría quedar asentado en un proyecto de historia oral. ¿Cuál de estos *World of Warcraft* debería ser preservado? La respuesta depende de los objetivos y la misión que tenga la institución interesada en su preservación. Antes de decidir cuál preservaría



usted, es importante que aclare el propósito de su preservación y que lo alinee con los planes para el desarrollo de la colección.

Lo anterior establece dos preguntas clave sobre el propósito de la preservación. En primer lugar, ¿qué aspectos de lo que deseamos preservar son significativos? En segundo lugar, ¿qué es necesario hacer para asegurarnos de que lo que nos importa persista en el futuro? Es sorprendente la frecuencia con la que estas interrogantes son omitidas. En muchos casos, las instituciones comienzan proyectos de preservación con respuestas tácitas o inarticuladas al respecto. Si usted no se ha tomado el tiempo de resolverlas, es poco probable que lo que ha adquirido, gestionado, descrito y a lo que ha proporcionado acceso cumplirá de manera adecuada con un objetivo que no fue establecido. La naturaleza desordenada de las plataformas anidadas de la información digital hace que esto sea un desafío. Además, lo que más les interesa a los usuarios es el significado que se puede obtener de las colecciones de los objetos, no de los objetos mismos. Así, es crucial ser deliberadamente claros en cómo respondemos a estas preguntas en un contexto dado. Es por eso que la preservación digital siempre debe estar basada en la articulación del propósito de la preservación.

Una declaración formal del propósito de la preservación establece con precisión por qué se seleccionó un conjunto de objetos digitales y qué características estructurales deben ser atendidas para que su contenido pueda usarse de acuerdo con el objetivo para el que fue recopilado. El establecimiento del propósito de la preservación comienza al responder la pregunta “¿por qué estamos haciendo esto?”. Articularlo da como resultado un diálogo constante entre lo que importa de las cosas que usted desea preservar, sus políticas de preservación digital y su enfoque continuo en el desarrollo de colecciones. En muchas organizaciones, estas diferentes líneas de trabajo se consideran áreas de actividad en principio distintas. Sin embargo, si no hemos dejado en claro los propósitos de la preservación de un conjunto particular de contenidos, entonces no contaremos con la base para asegurarnos de que las políticas puedan respaldar su objetivo. Del mismo modo, si la planificación del desarrollo de una colección digital ocurre fuera de las discusiones sobre el propósito de la preservación, entonces una institución no contará con los elementos para asegurar los recursos necesarios para llevar a cabo sus planes y deseos. Es necesario ser decidido y estratégico, y para ello es vital articular de forma explícita el propósito de la preservación. Esto es fundamental, dado que decidir qué es lo más importante sobre un conjunto de materiales puede conducir a enfoques en extremo diferentes al momento de recopilar, preservar, organizar, describir y proporcionar el acceso a los contenidos.



En este capítulo mostraré una serie de ejemplos de colecciones digitales y objetos en los que algunas instituciones dedicadas al patrimonio cultural han trabajado para recopilar y preservar. Primero, me enfocaré en las dificultades que existen para definir qué es lo que importa: debido a sus contextos de creación y sus diferencias en cuanto a lo que puede ser valioso sobre ellos, hay objetos similares que pueden requerir diferentes enfoques para su preservación. Después, exploraré hasta qué punto lo importante yace a veces fuera de lo que podrían parecer los límites predeterminados de un objeto dado. Esto abre una discusión más amplia sobre cómo las posibilidades de la recopilación digital pueden apuntar a algunas estrategias para el desarrollo de una colección. Como en cualquier oficio, los principios y enfoques surgen a través de la praxis. En conjunto, los ejemplos subrayan cómo un propósito de preservación bien articulado hace que las colecciones resultantes sean más fáciles de evaluar y más transparentes para sus futuros usuarios.

A lo largo del capítulo utilizaré estos ejemplos para mostrar cómo, a partir de la misión curatorial de una institución, usted podría tomar diferentes decisiones sobre la importancia o el significado de un objeto determinado. Por lo tanto, cada caso resalta cuestiones clave para la articulación del propósito de la preservación. Lo ideal sería que las instituciones establecieran procesos para crear un documento con los objetivos de la preservación. Es posible que éstos se presenten de varias maneras; una de ellas puede ser simplemente un breve informe escrito sobre lo que se considera significativo de un conjunto de contenidos, acordado y comprendido por las partes interesadas. Como demuestran varios de los ejemplos, este tipo de objetivos deben examinarse y revisarse durante la práctica, en sí, de adquirir, organizar, describir y proporcionar acceso al contenido. Si usted ya cuenta con una gran cantidad de contenido digital en un repositorio para el que nadie ha articulado todavía el propósito de su preservación, por favor, retroceda y hágalo.

Identificar lo que es importante en el contexto

Para poder articular el propósito de la preservación, usted necesita identificar qué es lo que importa del contenido de acuerdo con la misión de su institución. En algún momento, muchas personas en el campo de la preservación digital trataron de identificar nociones de las propiedades significativas de varios tipos de archivos y formatos digitales que fueran independientes del contexto. Si bien este enfoque tiene su mérito, se ha abandonado, en gran medida, ante el descubrimiento de que lo relevante en esencia depende del



contexto.⁵⁵ Es decir, la importancia no es una propiedad inherente a las formas de contenido, sino una característica emergente del contexto en el que éste se creó y el propósito al que sirve en la colección de una institución. Incluso cuestiones básicas, como qué registros u objetos son idénticos o lo que constituye un original y una copia, contienen una serie de convenciones que asumimos sobre el criterio que determina qué es lo importante de la naturaleza de estas cosas.⁵⁶ Comprenderlo es una parte clave del por qué las organizaciones han comenzado a cambiar a un enfoque que establece con anterioridad los objetivos de la preservación.⁵⁷

Discutir algunos ejemplos de diferentes tipos de contenido digital ayudará a ilustrar lo difícil que puede ser identificar y articular su significado. Así que consideraremos brevemente la importancia del contenido de la *laptop* de Salman Rushdie, los disquetes de Carl Sagan y los CD y DVD enviados al Veterans History Project. Los tres tienen una gran cantidad de información que puede resultar relevante. Al comprender el contexto en el que se crearon y el propósito para el que se han recopilado, pueden surgir temas para establecer los objetivos de la preservación.

La laptop de Salman Rushdie

Los archivos del autor Salman Rushdie fueron adquiridos por la Rose Manuscript, Archives, and Rare Book Library de la Emory University. Esta colección se entregó con tres computadoras portátiles que contenían correspondencia por correo electrónico, borradores de escritos y otras notas; exactamente el tipo de material que busca un documentalista para crear el archivo personal de una figura literaria. Es una gran colección.

Hasta cierto punto, es sencillo copiar los archivos de la computadora. Sin embargo, dado que se trataba del entorno de trabajo digital personal de una importante figura literaria, el equipo que trabajó con las *laptops* se dio cuenta

55 Angela Dappert y Adam Farquhar, “Significance Is in the Eye of the Stakeholder”, en *Research and Advanced Technology for Digital Libraries*, ed. de Maristella Agosti *et al.* (Berlín; Nueva York: Heidelberg / Springer, 2009), 297-308, http://link.springer.com/chapter/10.1007/978-3-642-04346-8_29.

56 Geoffrey Yeo, “‘Nothing Is the Same as Something Else’: Significant Properties and Notions of Identity and Originality”, *Archival Science* 10, núm. 2 (2010): 85-116.

57 Colin Web, David Pearson y Paul Koerben, “‘Oh, You Wanted Us to Preserve That?!’ Statements of Preservation Intent for the National Library of Australia’s Digital Collections”, *D-Lib Magazine* 19, núms. 1-2 (2013), DOI: 10.1045/january2013-webb.



de que gran parte del otro tipo de información que existe en la computadora (sus interfaces y aplicaciones) era de interés potencial para los investigadores.⁵⁸ Como ejemplo, Rushdie hizo un amplio uso de la funcionalidad Sticky Notes del sistema operativo Mac, que le permitía mantener notas codificadas por colores similares a un *post-it*.

Dado el contexto de las *laptops* y su importancia en relación con el propósito de la biblioteca, que era preservar los documentos personales de Rushdie, tenía sentido desarrollar un plan para mantener una copia del contenido de la computadora (después de eliminar los archivos que no eran de interés para la colección) y seguir después una estrategia de emulación del entorno de escritura para que los investigadores pudieran explorar cómo lo veía el autor. Ahora se puede acceder a una de esas computadoras portátiles y hacer lo mismo. Revisaré el tema de la emulación y la virtualización en el próximo capítulo y una vez más en el capítulo sobre acceso, pero, en pocas palabras, diré que éstas son técnicas que permiten proporcionar acceso al contenido digital a través de una réplica o una manifestación virtualizada de su *software* y su sistema operativo original.

En este caso, era razonable establecer un propósito de conservación ambicioso para las *laptops*, y, en efecto, los archivistas lograron un grado bastante elevado de representación auténtica. Dado que estas computadoras fueron utilizadas por una figura literaria importante, era prudente invertir tiempo y recursos para tratar de obtener una experiencia lo más rica y genuina posible de su información y sus archivos digitales. Las cualidades artefactuales del contenido en estas máquinas tienen el potencial de ser valiosas por la razón principal por la que se adquiere una colección como esta. Sin embargo, diferentes contextos de uso para los medios digitales pueden sugerir distintos enfoques, como se verá en los siguientes ejemplos.

Los disquetes de Carl Sagan

Los documentos del astrónomo Carl Sagan fueron adquiridos por la Library of Congress Manuscript Division. La colección masiva de documentos personales llegó con una cantidad significativa de contenido nativo digital. La colección incluye 379 disquetes que contienen más de 18 mil archivos, la mayor parte de ellos corresponden a documentos de WordPerfect. A partir de esa descripción,

58 Para obtener más información sobre el enfoque del procesamiento y el acceso a los documentos y las computadoras portátiles de Rushdie, ver Laura Carroll *et al.*, “A Comprehensive Approach to Born-Digital Archives”, *Archivaria* 72 (2011): 61-92.



podría parecer que este contenido es un gran candidato para hacer copias forenses (bit por bit) de cada disco. A primera vista, parece que este caso podría ser muy similar al de la *laptop* de Rushdie. Es decir, dado que se trata de una colección de documentos personales, parece lógico tratar de obtener las copias más fieles de cada disco para que permitan recuperar archivos eliminados o encontrar información oculta en las capas presentes en ellos. Sin embargo, comprender el contexto detrás de estos discos nos lleva a un enfoque diferente.

Carl Sagan no era un gran usuario de computadoras, lo que puede parecer extraño dado que sus documentos personales contenían cientos de disquetes. El proceso de escritura de Sagan involucró una constelación de tecnologías y personas. Cuando trabajaba en libros, en los años 80 y 90, dictaba sus secciones en una grabadora,⁵⁹ y luego las enviaba para que se transcribieran. La mayor parte de la información contenida en los disquetes está conformada por las transcripciones de esos audios. A partir de ahí, Sagan marcaría las impresiones de las transcripciones y las editaría, revisaría y ensamblaría en borradores. Por lo tanto, el archivo Sagan está, en gran medida, lleno de impresiones de los archivos de esos disquetes. Las impresiones son interesantes en sí mismas porque tienen todas las ediciones y reescrituras manuales. Entonces, ¿qué es lo que importa sobre los archivos y los discos?

En este contexto, tiene sentido tener un propósito de preservación diferente para estos discos que para las *laptops* de Rushdie. Si bien alguien podría estar interesado en las peculiaridades de los disquetes de Sagan o sus archivos, en realidad no le van a decir mucho sobre el autor y sus prácticas de escritura. Como otro punto de comparación, a diferencia de los discos del dramaturgo Jonathan Larson (del capítulo 2), que incluían información oculta en los archivos de guardado rápido de la computadora que utilizó, los de Sagan ni siquiera fueron empleados directamente por él. De nueva cuenta, encontramos una distinción clave para establecer el propósito de la preservación. No es que los disquetes en general tengan algún conjunto de propiedades significativas o, para el caso, que los tipos de archivos en esos discos sí lo hagan. Lo que importa es el contexto en el que se usaron y la conexión entre ese contexto y el propósito de la adquisición de una colección. En el caso de los disquetes de

59 Para conocer más sobre las prácticas de escritura de Sagan, ver Trevor Owens, “Sagan’s Thinking and Writing Process”, en *Finding Our Place in the Cosmos: From Galileo to Sagan and Beyond* (Washington DC: Library of Congress, 2014), <https://www.loc.gov/collections/finding-our-place-in-the-cosmos-with-carl-sagan/articles-and-essays/carl-sagan-and-the-tradition-of-science/sagans-thinking-and-writing-process/>.



Larson, lo lógico es tener una expectativa más alta que con los de Sagan. Sin embargo, debido al papel que jugó la informática en su proceso de escritura, estos discos reproducen en gran medida el contenido que está disponible en el resto de la colección, por lo que parece lógico adoptar un enfoque mucho menos exigente y de alto contacto para capturar su información. Con sólo obtener una copia de todos los archivos y retenerla, es probable que sea suficiente. Incluso se podría argumentar que, dado que casi todo este material se reproduce en forma impresa en toda la colección, podría no molestarse en guardar los discos en absoluto.

Sin embargo, la información en los discos viene con varias ventajas importantes. Primero, los archivos ilustran y explican parte del proceso de escritura de Sagan. Por lo tanto, incluso si contienen información en gran medida redundante para el material que se imprime en el archivo, también proporcionan información adicional sobre el contexto para estudiar su proceso de escritura. En segundo lugar, los archivos de WordPerfect, que constituyen la mayor parte del contenido, están formados por texto de búsqueda y marcas de fecha y hora que permiten a los interesados buscar y estudiar la colección más fácilmente. Esto apunta hacia una cuestión que aparecerá más adelante, cuando se hable sobre el acceso: cada vez más, un conjunto de objetos puede servir como una forma de metadatos para otro conjunto. No es necesario digitalizar y ejecutar el reconocimiento óptico de caracteres en buena parte de esta colección, pues es posible utilizar el texto de los archivos y las fechas para realizar búsquedas de texto completo a través de gran parte de su contenido.⁶⁰ Esto da lugar a una considerable plusvalía de los archivos de WordPerfect sobre los registros en papel y las cintas de audio del archivo. Si bien el contenido puede ser redundante, la indexabilidad de los archivos digitales significa que tienen un potencial considerable en términos de acceso.

El punto clave de todo esto es comprender que el contexto para el uso y la creación de medios digitales puede sugerir diferentes enfoques de lo que es significativo sobre un conjunto dado de objetos. Los archivos en la *laptop* de Rushdie son el resultado del uso personal de su equipo informático. Los archivos de computadora de Sagan son mucho menos personales en ese

60 Para una exploración sobre los tipos de modos de acceso y descubrimiento que serían posibles utilizando la información de texto completo extraída de los archivos, ver Owens, “Mecha-Archivists: Envisioning the Role of Software in the Future of Archives”, *Trevor Owens: User Centered Digital History*, 27 de mayo de 2014, <http://www.trevorowens.org/2014/05/mecha-archivists-envisioning-the-role-of-software-in-the-future-of-archives/>.



sentido. De acuerdo con estos contextos de uso, el archivo de Rushdie tiene un valor estético que invita a prestarle más atención y cuidado que al archivo de Sagan. La importancia de un contenido en particular sólo puede aparecer y evaluarse en el contexto de su uso original y el propósito de preservación de la organización que está trabajando para preservarlo.

Los CD y DVD del Veterans History Project

El Veterans History Project (VHP), que forma parte del American Folklife Archive en la Library of Congress, recopila y preserva las experiencias de primera mano de los veteranos de los Estados Unidos. Contiene una variedad de materiales personales, entre los que se incluyen fotografías, cartas, diarios y mapas, pero la mayor parte de la colección se conforma de historias orales, muchas de las cuales ahora se transfieren al archivo en CD y DVD. Mientras que en algunos contextos, los CD o los DVD pueden considerarse como un objeto, un registro o un artefacto, en este caso, se piensa en ellos como portadores.⁶¹ Son un medio para transferir las historias orales y no sólo simples artefactos.

La distinción entre estos medios como portadores o artefactos puede parecer pedante, pero las diversas perspectivas y objetivos sobre ellos producen diferencias clave en la manera en que se aborda la recopilación y la preservación de contenido. Desde esta óptica, tiene sentido centrarse más en el contenido audiovisual que es transmitido en el medio y mucho menos en cuestiones relacionadas con los formatos o con idiosincrasias de la forma en que la información se almacena o se presenta en ese medio.

Si bien uno podría crear imágenes de disco de cada medio, en este caso, en verdad sería justificable sólo extraer los archivos de audio y video, y luego normalizarlos de forma consistente para crear un conjunto coherente de contenido al que sería más fácil garantizar el acceso a largo plazo. Este enfoque tendría como consecuencia la pérdida de información. Podemos imaginar a algún futuro usuario que podría estar interesado en la información almacenada que no sean los archivos de A/V o en cómo se ordenaron y colocaron en los discos. Por ejemplo, algunos de ellos incluyen menús de DVD para navegar por el contenido. Sin embargo, dado que el propósito de este proyecto es

61 Para más detalles sobre la manera en que el VHP procesa las colecciones nativas digitales, ver Owens, "Personal Stories, Storage Media, and Veterans History: An Interview with Andrew Cassidy-Amstutz", *The Signal* (blog), 6 de marzo de 2014, <http://blogs.loc.gov/thesignal/2014/03/personal-stories-storage-media-and-veterans-history-an-interview-with-andrew-cassidy-amstutz/>.



conservar los registros individuales de los veteranos y recopilar y preservar sus grabaciones, es totalmente razonable centrarse en el contenido de los audios y no en los elementos de los medios que los transportan. Así es como se establece el propósito de la preservación: hay que aclarar cuál es la intención de una iniciativa de recopilación y trazar los objetivos a través de las características y posibilidades de su contenido.

Al centrarse sólo en el contenido de los archivos de audio y video, es posible establecer un flujo de trabajo mediante el cual se normalice cada grabación. La colección resultante estaría mucho más estandarizada, y sería más fácil trabajar con ella. En el futuro, uno podría imaginar de qué manera un proyecto como éste podría posicionarse incluso mejor para generar ese nivel de consistencia al desarrollar una aplicación móvil que permitiera a las personas crear grabaciones, describirlas, etiquetarlas y subirlas de inmediato. Al tomar control de esta parte del proceso de manera más directa, el contenido sería aún más uniforme.

Este caso ilustra la forma en que un determinado propósito de preservación podría establecer un marco diferente para pensar en los medios digitales. Es decir, dado el objetivo de este proyecto, tiene mucho sentido centrarse menos en las cualidades artefactuales de los medios digitales y tratarlos como simples portadores de información audiovisual. Vale la pena señalar que, cuando un proyecto puede definir de manera firme su propósito, es posible producir un contenido normalizado, que será mucho más fácil de asegurar en el futuro. Éste suele ser el caso cuando una institución ejecuta un proyecto de documentación o un esfuerzo de digitalización. En los ejemplos de las colecciones de documentos personales de Sagan, Rushdie y Larson, una institución recopila y se queda con los archivos y formatos que el individuo usó. Mantener esos formatos originales abre algunos desafíos potenciales para la capacidad de representación futura del contenido. En cambio, cuando uno puede estandarizar la normalización, como en el caso de los DVD y CD del VHP, se logrará un conjunto de recursos más consistentes y con los que será más fácil trabajar. Al controlar los contextos y formatos que se utilizan o al hacer que la normalización sea parte del proceso de adquisición, es inevitable que las colecciones sean más fáciles de planificar y gestionar en el futuro. Esto ilustra cómo, en el diagrama del último capítulo, las flechas van del propósito de la preservación al trabajo de administración de formatos, disposición, descripción y acceso. Las decisiones tomadas al definir el propósito de preservación tienen ramificaciones claras y directas en el trabajo que se realiza en cada una de estas áreas.



A veces lo importante se encuentra fuera del objeto

Cuando se toma en serio el propósito de la preservación, parece que lo más significativo de un objeto en un contexto dado no siempre se registra y preserva mejor en él mismo. Es importante recordar que, en su mayor parte, hoy en día y en el futuro las personas están menos interesadas en las cosas en sí mismas que en su significado. Este último está ligado a una variedad de situaciones contextuales alrededor de los roles sociales, culturales y económicos que juegan los objetos en nuestras vidas. Así, podrían documentarse grabaciones de un objeto mientras se usa en su contexto original para demostrar de mejor manera su importancia. Ésta es una de las razones clave por las que es esencial contar desde el principio con propósitos establecidos y siempre regresar a ellos. Si bien se podrían guardar “las cosas”, es completamente posible que lo que importa no esté en su interior, sino que sea algo que pueda grabarse o documentarse mejor a través de otros medios. Como se señaló en el caso de los CD y DVD del VHP, cuando se produce este tipo de documentación, se suele ganar el control sobre los formatos, lo que facilita mantener el acceso al contenido.

Este punto es cierto para la preservación en general. Por ejemplo, usted podría conservar la maquinaria pesada de una fábrica, pero eso no le aportaría mucha información sobre la experiencia cotidiana de trabajar allí. Si lo que desea es conocer la vida de un obrero, es probable que un proyecto de historia oral con los trabajadores de la fábrica le brinde un mejor material para preservar. De la misma manera, usted también podría preservar las cartas y los artefactos que quedan en el Vietnam War Memorial. Sin embargo, las fotos y videos de cómo interactúan las personas con ese monumento y las historias personales que comparten allí, le darían diferentes tipos de comprensión sobre cómo funciona el monumento. Asimismo, usted podría preservar la partitura de una sinfonía, pero una grabación capturaría mejor la experiencia de estar en la audiencia. Entonces, a cierto nivel, hay una continuidad entre la preservación de los objetos digitales y los analógicos. Dicho esto, la naturaleza de las plataformas de los objetos digitales revela varias razones y posibilidades adicionales para la documentación como un enfoque de preservación. En todos estos contextos, comenzamos a movernos un poco fuera de los límites del propósito de la preservación hacia cuestiones más amplias del desarrollo de colecciones. En vez de definir la importancia de un objeto, como en los ejemplos anteriores, en estos contextos, lo dejamos de lado para preguntarnos qué objetos deberían ser seleccionados para su preservación.

Una serie de ejemplos ayudarán a ilustrar estos puntos. A continuación, describo de manera breve los enfoques para formar una colección de una obra



de arnet, una aplicación para una base de datos médica, un juego multijugador masivo en línea y una aplicación para iPad. Cada caso demuestra cómo, dependiendo del contexto en el que opera un objeto, podría tener más sentido recopilar y preservar el contenido que documenta los aspectos de su uso y su función que preservar el objeto en sí mismo.

Form Art: *¿Se trata del objeto o de cómo se ve?*

Form Art de Alexei Shulgin es un sitio web interactivo compuesto por elementos tomados de formulario básicos en HTML, es decir, cosas como casillas de verificación, botones de radio, cuadros de diálogo, etc. Creado en 1997, el trabajo primero se representó y fue visualizado en un navegador como Netscape 3. Con esto en mente, incluso entonces, la manera en que se veía en la pantalla dependía del navegador que se utilizaba. En otras palabras, en su contexto original, el artista ya había perdido un control significativo sobre cómo se visualizaría exactamente. Rhizome, una biblioteca digital vinculada al New Museum de la ciudad de Nueva York, conserva *Form Art* como parte de su ArtBase.⁶²

Esta clase de elementos de los formularios están diseñados para adoptar un aspecto distinto en diferentes navegadores. A continuación se puede ver un ejemplo de cómo el trabajo cambia su apariencia con el tiempo cuando uno lo ve en un navegador contemporáneo.

La imagen ilustra cómo en un nuevo navegador y un nuevo sistema operativo el trabajo subyacente puede parecer muy diferente. Todavía es reconocible, pero los gradientes y el color agregados en el navegador de Chrome de Mac ofrecen una representación visual muy diferente del mismo HTML subyacente.⁶³

A estas alturas la pregunta es qué importa más: la manera en la que se despliega en la pantalla, el HTML del sitio o ambos. Dado que ésta es una obra de arte y que su aspecto es una parte clave para comprender su estética, parece claro que lo que se muestra en la pantalla probablemente sea lo más importante. Al mismo tiempo, mantener el HTML sería igual de significativo para alguien que quiere poder desentrañar a detalle la forma en que se construyó y se diseñó.

62 Alexei Shulgin, *Form Art* [1997], Rhizome, acceso el 28 de abril de 2021, <http://rhizome.org/art/artbase/artwork/form-art/>.

63 Ben Fino-Radin, “Rhizome ArtBase” y “The Web Browser as Aesthetic Framework: Why Digital Art Today Looks Different”, *Vice*, 8 de mayo de 2012, <https://www.vice.com/en/article/z4yagw/digart-the-web-browser-as-aesthetic-framework-why-digital-art-today-looks-different>.

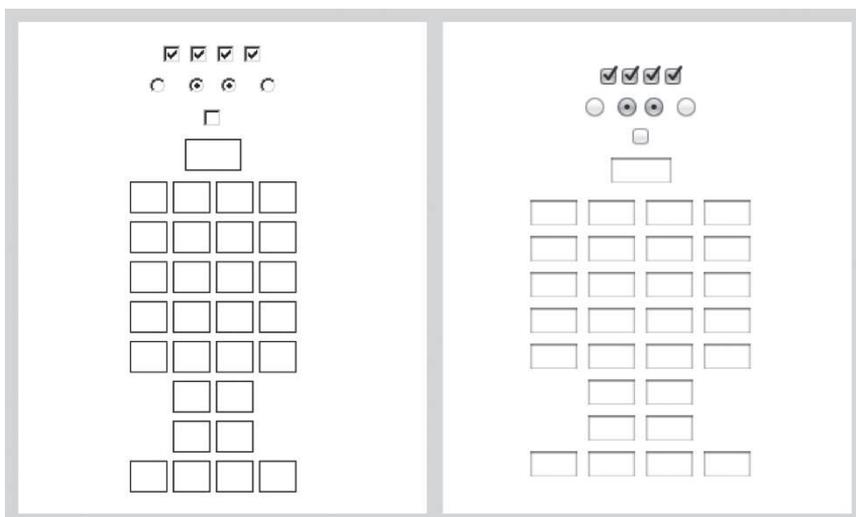


Figura 4. Capturas de pantalla de *Form Art* de Shulgin (1997), tal como se representan en Netscape 3 (izquierda) y en una versión reciente de Google Chrome (derecha). Cortesía de Ben Fino-Radin, <http://rhizome.org/art/artbase/artwork/form-art/>.

Como es evidente en la figura anterior, todavía es posible grabar y mostrar cómo se veía exactamente este trabajo, mediante una captura de pantalla. Ésta es una copia perfecta en píxeles de lo que apareció en el monitor en un formato de imagen estándar que será fácil de representar tanto ahora como en el futuro. Asimismo, conservar y mantener un navegador y un sistema operativo antiguos también es una forma viable para lograr este tipo de representación auténtica. Dicho esto, es inevitable que enfocarse en virtualizar o emular sistemas pasados en un *hardware* nuevo sea un proceso complejo, mientras que el tener al menos una imagen de referencia de cómo lucía en un momento particular en el tiempo es, con mucha seguridad, un recurso valioso para evaluar el nivel de autenticidad de sus siguientes representaciones. Aquí surge un principio general. La documentación (como la captura de pantalla) puede ser útil tanto como medio para preservar el significado como para crear material de referencia para triangular la importancia de un objeto o trabajo. En casos como éste, donde la documentación que se obtiene es trivial en tamaño, tiene mucha lógica incorporarla sin reservas como estrategia.

Vale la pena señalar que la manera en que se reproduce algo no es sólo una cuestión de *software*. Los diferentes procesadores de velocidad reproducen y representan un objeto de manera diferente. Las diversas tarjetas de sonido tocan los audios de formas distintas y variadas. Muchas de estas diferencias



sutiles son evidentes para los usuarios.⁶⁴ Además, los trabajos se ven diferentes cuando se procesan en un monitor CRT antiguo o en una nueva pantalla LCD. Es importante destacar que las diferencias entre uno y otro no se registrarían en una captura de pantalla, ya que son propiedades de las pantallas en sí y no de cómo se representan en ellas las imágenes. Aún así, existen maneras de lograr la apariencia de los monitores CRT. Para algunas partes interesadas, esa estética es lo suficientemente significativa como para garantizar el desarrollo de filtros que hagan que una pantalla LCD parezca un monitor CRT.⁶⁵ En suma, incluso registrar lo que una computadora está mostrando a través de capturas o videos de su propia pantalla, puede hacer que se pierdan aspectos de un trabajo que podrían ser significativos. Con ese fin, algunas personas que tratan con objetos nativos digitales han utilizado equipos de grabación de video y fotografía que les permiten documentar cómo se ven y suenan estos trabajos.⁶⁶ El punto aquí es que cuando los detalles sobre cómo se procesa una obra tienen particular importancia, siempre existe la opción de seguir adelante y grabarlo, ya sea a través del entorno informático o de manera externa con una cámara o un dispositivo de grabación de audio y video. Al final, se tendrá algo que no captura la interactividad del trabajo, pero mantiene una documentación estable y fácil de representar de cómo era exactamente en un contexto histórico particular.

Grateful Med: ¿Cuáles son sus límites?

Grateful Med es un programa de *software* desarrollado por la National Library of Medicine (NLM) para proporcionar acceso a una serie de bases de datos médicas. Se creó en 1986 y continuó en uso y evolución hasta 2001. Reconociendo su importancia, en 2016, la NLM incorporó a Nicole Contaxis como residente de administración digital para planificar su preservación.⁶⁷

Es fácil referirse a Grateful Med como si fuera una sola cosa coherente. Sin embargo, si se inspecciona con más profundidad, es casi imposible analizar sus

64 Margaret Hedstrom *et al.*, “‘The Old Version Flickers More’: Digital Preservation from the User’s Perspective”, *American Archivist* 69, núm. 1 (2006): 159-187.

65 Zach Whalen, “Close Enough” (presentación, MLA, 8 de enero de 2012), <http://www.zachwhalen.net/posts/mla-2012-close-enough/>.

66 Fino-Radin, “Take a Picture, It’ll Last Longer...”, *Ben Fino-Radin’s Blog*, 28 de agosto de 2012, <http://notepad.benfinoradin.info/2012/08/28/take-a-picture/>. [Nuevo blog <https://finoradin.info>].

67 Nicole Contaxis, “Grateful Med: Personal Computing and User-Friendly Design”, *Circulating Now* (blog), 28 de abril de 2016, <https://circulatingnow.nlm.nih.gov/2016/04/28/grateful-med-personal-computing-and-user-friendly-design/>.



límites. Como ocurre con cualquier *software* que haya estado en uso durante más de una década y media, tiene innumerables versiones. Dicho esto, si ése fuera el único problema, uno podría establecer la identificación de las versiones y luego obtener copias de cada una. Sin embargo, si usted iniciara cualquiera de esas aplicaciones, se daría cuenta muy rápido de que, en realidad, son sólo una interfaz para visualizar las bases de datos de la NLM. Entonces, cuando pensamos en Grateful Med, no sólo nos viene a la mente la interfaz, sino también lo que interactúa con ella. Si se buscara conseguir eso, sería necesario preservar las aplicaciones del cliente junto con las instancias del servidor *back-end*, es decir, del modo administrador, con las que se conectaron. Dado el largo periodo en que se ejecutó el *software*, esos sistemas *back-end* requerirían algunas constelaciones de tecnologías de *software* y *hardware* y sus dependencias. Y si esto no fuera suficiente, el contenido real, los datos a los que se accede a través de esa interfaz y ese servidor cambiaron y evolucionaron a través del tiempo. Entonces, si usted quisiera capturar la experiencia de utilizar el *software* en 1989, también necesitaría tener una copia de lo que estaba en las bases de datos de ese año. En resumen, como muchos sistemas digitales que nos interesan, Grateful Med es un gran cúmulo interconectado de aplicaciones, sistemas y conjuntos de datos. Cuantas más preguntas le haga, más grande, complejo y desalentador parece ser. ¿Qué esperanza hay para preservarlo?

Ésta es exactamente la razón por la que es vital dar un paso atrás y preguntarnos por qué estamos haciendo esto y qué podemos hacer con los recursos disponibles. Gran parte de la historia de Grateful Med está documentada en los archivos de la NLM. Para eso es la documentación, tanto digital como analógica, para que pueda proporcionar acceso a gran parte de la historia del *software*. A la par, el resto del trabajo en la NLM se enfoca en mantener y conservar las bases de datos. Entonces, aunque puede ser interesante proporcionar acceso a la información histórica, dada la cantidad de trabajo que se requeriría, no está claro si existen suficientes razones para hacerlo.

En algún punto de la investigación de Contaxis, apareció un *software* relacionado que resultó ser un objetivo mucho más fácil para la preservación y que conservaba las ambiciones más amplias de preservar Grateful Med. Esta aplicación, llamada How To Grateful Med, es un *software* independiente mucho más simple, que fue creado como tutorial para usar Grateful Med, e incluye búsquedas de muestra, por lo que, a este respecto, es una parte autónoma de esa infraestructura desperdigada. Aún mejor, fue creada como un medio para explicar y mostrar cómo usar el *software*. Por lo tanto, esta aplicación proporciona orientación e información sobre el funcionamiento del sistema que no estaría



presente en el *software* mismo. Con el fin de preservar una aplicación discreta que se puede reproducir en un entorno emulado o virtualizado y mostrarle a alguien cómo era el *software* y cómo funcionaba, la aplicación proporciona una meta ideal para la preservación. En este caso, parece que la mejor manera de preservar aspectos específicos de Grateful Med no fue intentar preservarlos, en absoluto, sino preservar algo que se creó para mostrar a las personas cómo usarlo. Al tomarse el tiempo para considerar el propósito de la preservación, observar qué otros tipos de información y documentación existen para respaldarlo y luego explorar varias metas potenciales de preservación, Contaxis pudo enfocarse en algo mucho más fácil con qué trabajar.

World of Warcraft: ¿se trata de una cosa o de una experiencia?

Como mencioné al principio del capítulo, *World of Warcraft*⁶⁸ involucra múltiples componentes de *software*. En realidad se parece mucho a Grateful Med. Cualquiera que lo juegue necesita una copia para ejecutarlo en su computadora, pero esas copias se actualizan y corrigen todo el tiempo. En consecuencia, el juego está en constante movimiento. Desde su estreno, ha habido innumerables lanzamientos adicionales de más misiones, lugares y otros elementos. Los aspectos de *World of Warcraft* cambian, a veces, hasta por semana. Junto con eso, el videojuego que se instala en la computadora no basta para jugarlo. Es necesario conectarse a través de la red a un servidor remoto que ejecuta un *software* adicional y que, a su vez, crea el mundo virtual en el que se convive con otros jugadores. Ese entorno también está experimentando ajustes y cambios constantes. Entonces, si usted quisiera capturar todo el ecosistema del *software* que constituye al juego, sería muy desafiante y difícil analizar qué es lo que más importa de él. Sin embargo, eso no necesariamente lo acercaría a lo que la mayoría de los jugadores o futuros académicos interesados en el juego piensen que es significativo.

Aun si usted pudiera mantener versiones del *software* y ejecutar su propia instancia del servidor cuando inicia sesión, sería difícil regresar a lo último que hizo en el juego. Entraría a un mundo gigante, solitario y vacío. Como cualquier videojuego multijugador masivo, toda la experiencia de *World of Warcraft* depende, en gran medida, de la interacción con miles de usuarios que se conectan al mismo tiempo. Para aclarar, ese mundo gigante, solitario y vacío es

68 Para una discusión más amplia sobre los retos en la preservación de mundos virtuales como *World of Warcraft* y *Second Life*, ver Jerome McDonough *et al.*, *Preserving Virtual Worlds. Final Report* (s. l.: Library of Congress, 2010), <http://hdl.handle.net/2142/17097>.



un videojuego, pero con relación a la manera en que debía ser experimentado, es decir, en cuanto a ser un lugar lleno de personas diferentes que interactúan y se relacionan entre sí, ese mundo no representa a *World of Warcraft*. De alguna manera, este tipo de objeto digital se parece más al escenario en el que se representa una obra teatral que a una pintura que cuelga en la pared. Como tal, tiene sentido que los enfoques para documentar una presentación en vivo funcionen mejor. También se puede argumentar que muchos de los aspectos más interesantes de un videojuego como éste son los tipos de cultura y los objetos que se producen a su alrededor, en cuyo caso lo más lógico sería abordarlo como un folclorista o un etnógrafo documentarían una comunidad.

Éste es exactamente el enfoque que han adoptado varias organizaciones interesadas en preservar *World of Warcraft*. El departamento de colecciones especiales de Stanford University ha recopilado un conjunto de videos generados por usuarios y compartidos en YouTube como una forma de capturar tanto algunos aspectos del mundo del juego como una parte del intercambio cultural y el discurso que se crea en torno al mismo.⁶⁹ En este caso, el artefacto (el videojuego en sí) puede ser sólo el trasfondo de lo que un usuario o un académico que estudia el juego podría ver como sus aspectos más significativos. Por lo tanto, la documentación creada por los usuarios y los miembros de una comunidad determinada puede ser mucho más fácil de representar y mantener en el futuro, desde el punto de vista técnico, pero también puede resultar mucho más rica como material para comprender los aspectos culturales de un juego.

Es más, todos los diferentes aspectos de este juego también podrían ser viables como objetos de preservación para un sinnúmero de instituciones con distintas misiones. El videojuego incluye una cantidad significativa de obras de arte, por lo que una organización podría adquirirlo como un objeto estético, en cuyo caso podrían tomarse decisiones sobre qué elementos de su arte visual y de su audio son las más importantes. *World of Warcraft* tuvo avances significativos en el diseño de videojuegos, por lo que, para una entidad que lo integrara a una colección de diseño, cosas como su código fuente podrían ser de particular interés. Este videojuego es un producto comercial global de gran éxito, por lo que sus registros corporativos digitales originales podrían ser valiosos para una institución que conserva datos sobre la historia de los negocios, la tecnología y las empresas. El sistema de chat dentro del juego ha transmitido miles

69 Henry Lowood, "Memento Mundi: Are Virtual Worlds History?", *iPRES 2009. The Sixth International Conference on the Preservation of Digital Objects. Proceedings* (s. l.: California Digital Library, 2009), 121-128, <http://escholarship.org/uc/item/2gs3p6jx>.



de millones de comunicaciones informales entre los usuarios que podrían ser atractivas para lingüistas, sociólogos y folcloristas interesados en estudiar los hábitos lingüísticos de la población global de jugadores. Todos estos aspectos de *World of Warcraft* se pueden coleccionar y preservar. Solamente si se articula por completo el propósito de la preservación y se conecta con el desarrollo de la colección es que se pueden hacer planes para garantizar el acceso a cualquiera de ellos en el futuro.

Planetary: adquisición del proceso detrás del objeto

Planetary es una aplicación para iPhones y iPads que visualiza la colección de música de los usuarios como cuerpos celestes y actúa como una interfaz para reproducirla. Las canciones se representan con lunas, los álbumes con planetas y los artistas con soles. La aplicación es un ejemplo bastante interesante de diseño gráfico y de interacción, por lo que un equipo del Cooper Hewitt Design Museum contactó a sus creadores y desarrolló un plan para obtenerla.⁷⁰

Los desafíos para adquirir una aplicación de iOS son múltiples. Incluso mientras se conversaba con el creador de Planetary, salió una nueva versión del sistema operativo iOS para la mayoría de sus usuarios. Esto significaba que la aplicación ya era obsoleta y funcionalmente no ejecutable antes de que se adquiriera. Si el objetivo era mantener una copia para utilizarla en un iPad, sería un desafío. El iPad se podría haber mantenido con el sistema operativo anterior, pero las dependencias existentes en esa situación (el *hardware*, el sistema operativo y la aplicación) se degradarían con el tiempo, por lo que la posibilidad de emular el sistema operativo era desalentadora. Las aplicaciones de iOS están diseñadas para hacer uso de las ventajas un tanto únicas del *hardware*. En cualquier caso, volviendo al objetivo del enfoque de Cooper Hewitt, que consistió en crear una colección útil de materiales para informar sobre diseño, el equipo decidió tomar un camino fundamentalmente diferente.

El código fuente subyacente de la aplicación se desarrolló a través del tiempo con Git, un sistema de control de distribución de versiones. Esto significó que todo el historial del diseño del código existía en el sistema de control de versiones del desarrollador. Dado que el enfoque del Cooper Hewitt Design Museum para su colección era documentar y preservar el acceso al diseño, de-

70 Para una descripción más profunda del proyecto, ver Sebastian Chan y Aaron Cope, "Collecting the Present: Digital Code and Collections" (presentación, MW2014: Museums and the Web 2014, 2-5 de abril de 2014), <http://mw2014.museumsandtheweb.com/paper/collecting-the-present-digital-code-and-collections/>.



cidieron centrarse en “Planetary: the software system” en lugar de en “Planetary: the iPad application”. En este caso, fue bastante fácil adquirir y proporcionar acceso a la copia completa del código fuente, con su historial de revisiones, informes de errores y comentarios de código. Después de obtener este contenido, el equipo lanzó el código fuente en Github, bajo una licencia de *software* de código abierto para que cualquiera pudiera descargarlo, explorarlo y construir sobre él. En este caso, el código fuente de Planetary documenta y proporciona acceso a información que no habría estado disponible si el equipo se hubiera centrado en preservar una versión final de la aplicación. El código fuente sirve como una extensa documentación para su creación. Además, este código, como la mayoría de los de su tipo, consiste en caracteres de texto y gráficos fáciles de representar y leer en formatos estándar. En consecuencia, el código y los archivos serán mucho más fáciles de mantener en el futuro.

El desarrollo de las colecciones: ¿estamos enfocados en las cosas correctas?

En este capítulo comencé seleccionando algunos aspectos de información digital particular almacenada en objetos individuales (disquetes y *laptops*) y luego retrocedí para cuestionar hasta qué punto la documentación de los objetos podría ser mejor para capturar la importancia que existe en otros lugares (audio y videgrabaciones, código fuente). Vale la pena retroceder una vez más para preguntarnos de manera más amplia cómo deciden las instituciones lo que es significativo. En particular, es relevante para el proyecto de este libro reflexionar sobre qué posibilidades pueden proporcionar los medios digitales para cambiar algunas de las formas en que nos relacionamos con las comunidades en el proceso de selección y el desarrollo de colecciones.

Como señalé antes, los objetos, las obras y los registros son de gran interés, no por derecho propio, sino por las formas en que nos permiten contar historias sobre las personas. Las bibliotecas, los archivos y los museos se crean y mantienen como parte de las estructuras de poder de la sociedad. Como tales, si y cuando aspiran a ser neutrales, legitiman pasivamente las prioridades y las maneras de ver que son parte del *statu quo*.⁷¹ Ningún archivista o bibliotecario es neutral; sólo se trata de si éstos trabajan de forma consciente para identificar y responder a los prejuicios del *statu quo* y si se aseguran de que sus colecciones representen múltiples voces o si al no hacer la pregunta sólo lo reproducen.

71 Howard Zinn, “Secrecy, Archives, and the Public Interest”, *Midwestern Archivist* 2, núm. 2 (1977): 14-26.



Profundizaré en un ejemplo para explorar cómo los nuevos enfoques para la recopilación digital nativa están reformulando el proceso del desarrollo de colecciones e incluso cómo rechazan la idea de que el desarrollo de colecciones es en sí mismo el fin de dicho trabajo.

Documenting the Now (La documentación del ahora)

Documenting the Now es un esfuerzo para explorar la ética de la recopilación y preservación del contenido de las redes sociales y para desarrollar herramientas y enfoques dirigidas a la colección y preservación de ese contenido de manera que sus creadores tengan una voz sobre cómo se recopila, preserva y utiliza su información. El contenido de las redes sociales, como Twitter y Facebook, representa una forma cultural novedosa a través de la cual todo tipo de personas interactúan entre sí y con sus culturas. Así, ciertos tipos de historias e interacciones que antes no eran perceptibles ahora se están recuperando y preservando. A medida que el papel que juegan estas plataformas en la vida cotidiana crece, se vuelve cada vez más importante para las organizaciones del patrimonio cultural trabajar en la recopilación y preservación de su contenido.

Tras el asesinato del adolescente desarmado Michael Brown en Ferguson, en 2014, gran parte de la respuesta de los activistas se desarrolló y coordinó mediante Twitter. Un pequeño equipo de archivistas comenzó a recopilar los tuits que mencionaban “Ferguson”, y rápidamente acumularon más de 13 millones de ellos.⁷² Mientras que los tuits se rigen por las restricciones legales establecidas por Twitter, debido a las formas en que este tipo de información se puede utilizar para vigilar y atacar a los manifestantes, hay cuestiones éticas en juego acerca de cómo se recopila, conserva y accede a este contenido que son más importantes que las que rigen los términos y condiciones para trabajar con datos de Twitter. Como resultado de estas exploraciones, el equipo se ha unido para dar luz a Documenting the Now, una colaboración entre archivistas y activistas para desarrollar herramientas, flujos de trabajo y enfoques que tiene el objetivo de recopilar, preservar y proporcionar acceso al contenido de redes sociales y que interactúa con sus creadores y les brinda una voz.

Si bien es posible consumir todo tipo de contenido digital, también es fundamental que las organizaciones reflexionen en qué medida están replicando los viejos procesos paternalistas en los que las instituciones recopilan datos e

⁷² Para más información sobre este tema, ver Bergis Jules, “Documenting the Now: #Ferguson in the Archives”, *On Archivy*, 8 de abril 2015, <https://medium.com/on-archivy/documenting-the-now-ferguson-in-the-archives-adcbe1d5788>.



información sobre varios pueblos en lugar de hacerlo junto con ellos o apoyarlos para que ellos trabajen con su propia información. Las redes sociales están abriendo nuevos géneros de objetos que exigen la atención de las instituciones del patrimonio cultural y también provocan nuevos problemas éticos importantes en torno al trabajo de la preservación y el acceso. Cuando buscamos desarrollar estrategias para la recopilación y el establecimiento de los objetivos de la preservación, es importante que no sólo nos enfoquemos en los tipos de contenido que se han recopilado en el pasado, sino que también consideremos formas nuevas y emergentes de contenido digital que puedan ofrecer ventajas adicionales para documentar el presente.

Del propósito a la acción y de regreso

Los ejemplos compartidos en este capítulo ilustran cómo el enfoque adoptado para preservar cualquier conjunto de contenido digital variará de manera drástica según la naturaleza de los objetos y lo que sea importante de ellos para usted o para su institución, además de cómo quiere que las personas puedan acceder a él en el futuro. A primera vista, a menudo parece que la preservación de algún objeto o conjunto de objetos es sencilla. Sin embargo, cuando usted comienza a preguntarse por qué está preservando algo, con frecuencia se vuelve mucho más complejo y extraño. Cuando comienza a indagar en los propósitos de la preservación, se vuelve claro que esto se conecta a todos los aspectos del desarrollo de una colección y a la planificación estratégica de una institución.

La discusión sobre los propósitos de la preservación en estos ejemplos ya ha tocado algunos temas que se verán en los próximos capítulos. Las decisiones que uno toma sobre los propósitos tienen efectos de reacción en cadena sobre las otras áreas y las posibilidades de la información y los medios digitales. Las consideraciones éticas en cada uno de los capítulos por venir tienen el poder de obligarlo a retroceder y a redefinir sus propósitos. Volviendo al ejemplo de Grateful Med, es posible preservar la masiva entidad en expansión que son todos los diferentes sistemas y *software* con los que se interactúa, pero para hacerlo, se requerirían muchos más recursos de los que uno podría justificar para participar en la actividad. Éste es el quid del proceso para todos los interesados en la preservación digital. Debemos anclar nuestro trabajo en propósitos claros para que sea inevitable tomar decisiones sobre lo que es práctico, dada la naturaleza de nuestros objetos, nuestras obligaciones éticas y los recursos disponibles, y para que si necesitamos hacer concesiones sobre ellos, lo hagamos deliberadamente.

Capítulo 6

La gestión de copias y formatos

Una pequeña biblioteca rural administrada por un solo bibliotecario tiene una colección de fotos digitales. Una biblioteca universitaria con 15 empleados tiene una variedad de colecciones digitales, compuesta por materiales manuscritos nativos digitales, copias del sitio web del periódico estudiantil y archivos PDF de tesis de licenciatura, entre otros. Una biblioteca de una universidad grande con cien empleados tiene colecciones digitales heterogéneas de tamaño considerable, demasiado numerosas para mencionar en este párrafo.

Las diferencias significativas en cada uno de estos contextos, en términos de recursos, tiempo del personal y estructura organizacional, debería dar como resultado enfoques muy distintos en el trabajo central de la preservación digital: desarrollar un proceso para la administración de copias y la resolución de cuestiones relacionadas a la reproducción a largo plazo de formatos digitales. Este capítulo explora esas cuestiones y las realidades de la planificación para la preservación digital en distintos contextos.

La gestión de copias y formatos como parte del cuidado de las colecciones digitales

Para comenzar, será útil contextualizar las prácticas presentadas en este capítulo junto con otras prácticas reconocidas por mucho tiempo en el cuidado de colecciones analógicas. Bibliotecas, archivos y museos han desarrollado nociones relacionadas, pero diferentes, en el trabajo de preservación analógica. Éstas incluyen una variedad de categorías: conservación, restauración, reubicación y modificación de formato para la preservación. Hablando de manera general, la gestión y el cuidado de las colecciones se ha convertido en el enfoque central



del trabajo base para la preservación en diferentes instituciones culturales y patrimoniales.⁷³

Dentro del marco del cuidado de las colecciones, el principal objetivo es establecer condiciones generales y seguras para el almacenamiento y la gestión de los materiales. Éstas deben ser las que funcionen mejor para todos los distintos tipos de objetos que se deseen preservar ahora y en el futuro. Para objetos físicos, esto en principio se enfoca en proveer las condiciones ambientales que retarden el deterioro de los materiales: asegurarse de que las colecciones estén almacenadas en un espacio seco, a la temperatura adecuada, con un sistema adecuado de circulación de aire, y en donde haya pocas probabilidades de ser dañados por una inundación, en un espacio con un sistema diseñado apropiadamente para el control de incendios, etc. Este capítulo se trata de trasladar la idea general del cuidado de las colecciones al contexto de la preservación digital o a dos áreas principales de actividad: primero, el establecimiento de sistemas y prácticas para gestionar múltiples copias de objetos digitales en el entorno de almacenamiento que se haya establecido y, segundo, el desarrollarlo de planes generales para la mitigación de posibles problemas relacionados a los formatos digitales.

Siguiendo la sabiduría de las profesiones dedicadas a la preservación cultural, es fundamental establecer sistemas y prácticas universales que garanticen que las condiciones básicas y los entornos del almacenamiento son seguros para la preservación. En lo que respecta a la preservación digital, esto implica establecer sistemas, flujos de trabajo y prácticas que gestionen, primero y ante todo, la reproducción de la información digital. El trabajo básico de la preservación digital es hacer, administrar y revisar copias. Después de esto, tiene sentido considerar los otros aspectos que ocurrirán sólo para tipos particulares de objetos y archivos. No podemos predecir el futuro, pero debemos pensar de manera estratégica en los posibles riesgos que puedan darse en el futuro para el acceso a los formatos digitales.

Este capítulo se enfoca sobre todo en el establecimiento de los requisitos para un sistema de almacenamiento digital. Los capítulos restantes se centran en cuestiones vinculadas a la adquisición de las colecciones individuales. En contraste, el proceso central para la gestión de copias y políticas alrededor de los formatos que se describen aquí debería ser la infraestructura general base. Con esto en mente, su enfoque general central para la preservación digital con

73 Bearman, *Archival Methods*, capítulo 2.



respecto a las copias y formatos deberá ser algo que esté pensado para ajustarse a los contornos de la misión de su institución y sus recursos: se trata de la infraestructura digital base que usted establece para permitir un trabajo más individualizado y necesario para cada colección en particular.

La primera mitad de este capítulo se centra en explicar los elementos y necesidades fundamentales para realizar la gestión de copias. Hacer esto es lo que nos permite mitigar los riesgos de la pérdida de datos o lo que comúnmente se conoce como “preservación de bits”. En concreto, ilustro cómo las prácticas de preservación de bits deben aplicarse de forma distinta en instituciones culturales y del patrimonio de diferentes tamaños, así como en el trabajo con contenidos de distinto tipo. Éste no es un enfoque para sistemas de almacenamiento que se ajuste a todos los casos. Por eso, es vital que una organización tome decisiones adecuadas en el diseño de sistemas, con base en sus recursos, sus objetivos y la naturaleza del contenido con el que trabaja. Después de eso, doy un giro para considerar problemáticas referidas a los formatos y su reproducción auténtica. También discutiré algunos aspectos sobre esta última más adelante, en el capítulo sobre acceso multimodal. Sin embargo, en este capítulo, me concentraré en una serie de enfoques y estrategias que las instituciones culturales y del patrimonio utilizan para administrar los riesgos de inaccesibilidad al contenido en el futuro dentro de su trabajo de diseño y desarrollo de sistemas de almacenamiento y depósito.

La preservación de bits

La preservación de bits, asegurar copias auténticas de información digital al nivel de bits en el futuro, es posible a través del establecimiento de un sistema de almacenaje y la administración de copias múltiples. Para ello, se deben realizar tres procesos centrales: uno para gestionar sus copias múltiples de datos; otro para administrar y hacer uso de la fijeza de la información para asegurar que pueda dar cuenta de toda su información digital; y, por último, uno necesario para establecer prácticas y procesos que garanticen que sus datos están seguros y que no serán corrompidos, dañados o borrados de forma inadvertida o maliciosa. Si se tiene más recursos, hay mejores formas de sistematizar y automatizar estos procesos, pero para colecciones relativamente pequeñas es posible hacerlo con pocos recursos y tiempo limitado y tener la confianza de que se tienen copias auténticas, siempre y cuando alguien continúe preocupándose y poniendo atención al contenido.



El mantenimiento de copias múltiples

En el campo de la preservación digital, ponemos nuestra confianza en tener copias múltiples. No podemos fiarnos de la durabilidad de los medios digitales, por lo que debemos hacer copias múltiples. Por suerte, como lo dije antes, la creación de copias es hasta cierto punto trivial. Sin embargo, una cosa es hacer copias y otra por completo distinta es administrar la reproducción sistemática del contenido digital.

Aquí hay una diferencia clave entre la gestión de copias para la preservación digital y simplemente mantener un respaldo de los datos. En el segundo caso, todo lo que usted necesita es una copia completa del sistema que esté usando para poder recuperar los datos en caso de que se dañe. Sin embargo, cuando configure su sistema y el proceso para administrar copias para la preservación digital, necesita hacer más que eso. Es necesario que establezca prácticas para definir con precisión qué contenido va a preservar y que se asegure de que la organización del contenido será inteligible para los usuarios en el futuro. Éste es un asunto al que volveremos más adelante, en el capítulo sobre la organización y descripción de los contenidos digitales. Por ahora, es suficiente subrayar la importancia de establecer una separación entre los archivos de trabajo que se editan y reorganizan de forma regular y el contenido digital que se ha convertido en contenido estático y que necesita ser preservado. Ésta es una distinción fundamentalmente intelectual. A los distintos sistemas de almacenamiento que se establezcan no les importa si los archivos se encuentran activos o si se están archivando. Sin embargo, es importante que usted tenga claridad sobre esto, pues se conecta de manera directa con los puntos siguientes. Usted necesitará gestionar la información de acuerdo con la integridad del contenido y su seguridad, ambas dependerán de las reglas que establezca para el manejo de copias.

Entonces, ¿cuántas copias debe tener y dónde debe guardarlas? Hasta cierto punto, mientras más copias mejor. Los niveles de preservación digital de la National Digital Stewardship Alliance (NDSA) recomiendan al menos tres, y esto se ha convertido en el estándar para muchas organizaciones culturales y del patrimonio. Otros enfoques y sistemas incluyen el almacenamiento de seis o más copias. En vez de prescribir un número mágico, creo que lo mejor es entender con exactitud contra qué tipo de amenazas de pérdida de información nos estamos protegiendo. En principio, tener al menos dos copias ayudará a asegurarse de no tener un único lugar de riesgo. Una segunda copia es esencial. Si usted no tiene una segunda copia del contenido, pero le preocupa asegurar el acceso a él, deje de leer este libro ahora mismo y vaya a



hacer una. En serio. Después de eso, es importante verificar que no se están guardando ambas copias en el mismo lugar. Si tiene cuatro copias en discos duros diferentes en su oficina, pero ocurre un incendio en el edificio, perderá las cuatro copias. Entonces, una vez que tenga al menos dos copias, necesita establecer prácticas para mitigar las distintas amenazas de desastre. Eso puede ser tan simple como mantener una segunda copia en casa, intercambiar copias con otra organización al otro lado del país o utilizar un proveedor de almacenamiento como Amazon o DropBox para almacenar una de las copias.

Otra consideración importante es no mantener todas esas copias en el mismo tipo de *hardware*. Por ejemplo, los discos duros tienen una duración de vida específica, por lo que, si compra varios discos duros hoy para mantener copias múltiples, existe la posibilidad de que todos puedan fallar al mismo tiempo. En relación con este punto, algunos profesionales del área de la preservación digital recomiendan mantener copias en medios de distinto tipo. Usted puede tener copias en discos duros (medio magnético) y otras en CD (medio óptico). Esto le permite proteger su contenido de eventos como los “pulsos electromagnéticos” (EMP), que son capaces de dañar a los medios magnéticos, pero no de afectar a los medios ópticos. Considerando esto, cada vez más organizaciones están dejando de seguir la sugerencia de utilizar medios distintos, en parte porque, primero, no está claro qué tan significativa es la amenaza y, segundo, porque con el tiempo se ha vuelto más difícil hacer uso de otros tipos de medios a gran escala.

Tener varias copias es fantástico, pero en última instancia es inútil si no se tiene un proceso bien articulado para administrar esas copias. Hay plataformas de código abierto, como Lots of Copies Keep Stuff Safe (LOCKSS),⁷⁴ que están diseñadas para realizar este tipo de administración.⁷⁵ Este tipo de sistemas a menudo incluye funcionalidades para revisar y administrar la integridad y para tratar con la seguridad de la información, dos aspectos que discutiré a continuación.

74 N. del T.: el acrónimo en inglés LOCKSS se traduce al español como “muchas copias mantienen las cosas a salvo”.

75 Victoria Reich y David Rosenthal, “Distributed Digital Preservation: Lots of Copies Keep Stuff Safe”, en *Proceedings Indo-US Workshop on International Trends in Digital Preservation* (Pune: Human Centered Design & Computing C-CAD, 2009), 51-55, <https://web.archive.org/web/20160507064543/http://lockss.org/locksswiki/files/ReichIndiaFinal.pdf>.



La administración de la integridad

En la discusión sobre la naturaleza de los medios digitales, mencioné de manera breve la idea de integridad. Éste es un concepto un poco confuso, pero si hay un solo término que se necesita saber de la jerga de la preservación digital, es integridad. Es más fácil entender la integridad en el contexto de la pregunta “¿cómo saber si un objeto digital ha cambiado o se ha degradado?”. La forma de responder a esta interrogante es estableciendo la integridad o la estabilidad de un grupo de objetos digitales.⁷⁶ Lo que es un poco confuso sobre esto es que no decimos que el contenido está fijado a un medio en particular. Al contrario, al mantener el foco de la preservación digital en el contenido informativo del objeto, tratamos de confirmar si éste es idéntico a lo que era la última vez que fue revisado. Dado que no estamos hablando sobre artefactos, sino de información, se trata en realidad de si aún tiene la secuencia exacta de información que tenía antes. En este sentido, la integridad es un sinónimo del término “integridad al nivel de bits” que es un poco más descriptivo.

Entonces, más allá de mantener copias múltiples de sus objetos digitales, es necesario que usted se asegure de que está manteniendo la integridad de esos objetos. Al hacer copias o al transferirlas de un medio de almacenamiento a otro, es posible que algunos procesos de copiado fallen y que la información se pierda. También puede ocurrir, aunque es menos probable, que se pierda cuando los datos están en reposo. Esto usualmente se conoce como “bit rot” y es (en general) una amenaza menor al acceso a la información a largo plazo comparada con la pérdida de datos que puede ocurrir durante una transferencia. La información digital siempre está codificada en un medio físico y cuando este medio se degrada, con el paso del tiempo, la información binaria codificada en él puede volverse menos legible. Esto puede derivar en la pérdida de algunos bits o errores de lectura. En cualquier caso, la manera en que usted controla estos riesgos es a través de la recopilación y de la verificación de la información sobre la integridad (de los datos) de forma periódica y, cuando haya errores, realizando labores de reparación para asegurarse de que todas las copias que se mantienen están impecables.

Existen varios tipos de información sobre la integridad con los que usted puede trabajar. Al nivel más básico, puede verificar el tamaño o el número de

⁷⁶ Para más información sobre integridad, ver Paula de Stefan *et al.*, “Checking Your Digital Content: What Is Fixity, and When Should I Be Checking It?” (reporte de trabajo, NDSA, 2014), <http://hdl.loc.gov/loc.gdc/lcpub.2013655117.1>.



los archivos. Si el contenido digital que quiere preservar es de cien megabytes compuestos de 50 archivos, entonces, verificar si tienen el mismo número y el mismo tamaño da cierto nivel de certeza de que tiene los mismos materiales. Este tipo de método (tamaño y número de archivos) es útil; sin embargo, hasta cierto punto, es sencillo generar y administrar formas más poderosas de información sobre la integridad a través de la creación de funciones *hash* criptográficas MD5 o SHA-1.

Una función *hash* criptográfica es un algoritmo que toma un cierto conjunto de datos (por ejemplo un archivo) y crea una secuencia de caracteres que puede servir como una huella digital para ellos. Incluso cambiar un solo bit en un archivo generará una secuencia diferente en su totalidad. Por ejemplo, yo generé un *hash* MD5 para un archivo con el valor “4937A316849E472473608D4347EBE9EF”. Ahora, si hago el *hash* para otra copia de ese archivo y recibo el mismo resultado, puedo tener una certeza más o menos alta de que esas dos copias son exactamente iguales. De la misma forma, si tengo guardado el valor de la última vez que generé el *hash* para ese archivo, cuando vuelva a hacerlo y reciba ese mismo valor otra vez, me sentiré bastante seguro de que ese archivo no ha cambiado ni ha sido alterado. Para tener una idea de qué tan alta puede ser esa seguridad, vale la pena mencionar que, como caracterizador de individualización, el *hash* MD5 ofrece más confianza que un test de ADN para la identificación de una persona.⁷⁷ Ahora, la gente le dirá que el MD5 no funciona y que por eso debería utilizar otro algoritmo, como SHA-1 o SHA-256. El MD5 no funciona en el contexto de la criptografía, pero es preciso destacar que esto no es particularmente relevante en su uso para la identificación en la forma en que nosotros lo utilizamos. Dicho esto, las funciones SHA tienen niveles adicionales de seguridad, pues son (al menos por ahora) seguros en un sentido criptográfico. No ahondaré en lo que eso significa, pues va más allá del objetivo de este libro, pero si le interesa, hay una gran riqueza de información muy profunda sobre el tema en Wikipedia.

Para administrar la integridad del contenido, usted necesitará recopilar y verificar la información sobre la integridad y luego deberá usarla para reparar los datos. Esto genera una pregunta: ¿qué tan seguido se debe verificar la integridad del contenido? Otra vez, es importante primero entender en qué lugares hay mayor riesgo de pérdida. Usted tiene una mayor probabilidad de

⁷⁷ Warren G. Kruse y Jay G. Heiser, *Computer Forensics. Incident Response Essentials* (Boston: Addison-Wesley Professional Education, 2002), 89.



perder su información en la transferencia de datos, por lo que, en una situación ideal, siempre deberá tratar de comprobar la información sobre la integridad (de los datos) antes de transferir los materiales para poder compararlos después. Hay dos programas de código abierto muy buenos, Fixity y Exactly, que pueden realizar esta función si desea transmitir datos. De forma similar, existe una herramienta llamada Bagger, que organiza los datos siguiendo la especificación Bag-It. Los archivos “baggeados”, es decir, empaquetados, básicamente incluyen valores de *hash* para cada uno en un documento de texto titulado “manifiesto”, junto con la información que se transfiere o preserva. En este sentido, usted debe asegurarse de verificar la integridad del contenido al migrarlo a un nuevo medio de almacenamiento. Por esta razón, si mueve todo de un disco duro a otro, o de un sistema de archivo de cinta magnética a otro, es importante que constate que todo ha sido transferido de forma correcta.

Más allá de eso, si su sistema puede hacerlo, es buena idea verificar la integridad de todo su contenido en intervalos regulares de tiempo, por ejemplo, de manera mensual, trimestral o anual. Muchas organizaciones, en especial aquellas que utilizan sistemas de almacenamiento compuestos sólo de discos magnéticos giratorios (discos duros), realizan este tipo de verificación programada de forma regular. En contraste, las organizaciones que usan sistemas de almacenamiento basados en cintas magnéticas, por lo regular, no lo hacen. Una cinta magnética se degrada levemente cada vez que se lleva a cabo su lectura, por lo que es mejor tener algún grado de confianza en el propio medio de almacenamiento en lugar de gastarlo muy rápido. También vale la pena mencionar que algunos sistemas de archivo (como el ZFS que mencioné antes) y algunos sistemas de cintas tienen procesos de verificación de integridad y reparación incorporados. En esos casos, la verificación de integridad regular y programada es realizada por el sistema, pero sin que el usuario tenga conciencia de ello. De forma similar, la mayoría de los proveedores de almacenamiento en la nube realizan procesos sistemáticos de monitoreo y reparación de datos. Sin embargo, en ambos casos, no hay transparencia para el administrador del repositorio en relación a lo que se hace y no se hace con sus datos. Llevar a cabo la verificación de la integridad de su contenido y mantener un registro de ello le permitirá tomar decisiones basadas en la certeza de que conoce el estado preciso de ese contenido. De cualquier manera, lo más importante es que, primero, usted establezca prácticas para administrar la información sobre la integridad (de los datos) y, segundo, que asegure su verificación, como mínimo, cada vez que necesite mover los datos.



La seguridad de la información

Usted puede gestionar copias múltiples y tener un sistema fabuloso para administrar y verificar la fijeza de la información y utilizar esa información sobre la integridad para reparar objetos digitales si es necesario, y aún así podría perder absolutamente todo. Basta con que exista la posibilidad de que una persona ingrese al sistema y borre todas sus copias. Esto podría ser de forma maliciosa (alguien que ha sido despedido ingresa y borra los datos), pero es mucho más común que este tipo de pérdidas ocurran cuando alguien ejecuta una orden errónea por accidente y borra más de lo que pretendía. Para controlar estos riesgos es importante detenerse a pensar quién puede hacer qué en términos del borrado de los datos o en cuanto al acceso físico a los dispositivos en los que se almacenan.

Al nivel más básico, usted puede solucionar esto a través de algo tan simple como mantener una copia de sus datos en un disco duro externo en un cajón cerrado con llave o en una caja fuerte. Si se puede acceder a todo su contenido a través de sistemas en línea, es muy probable que necesite, como mínimo, asegurarse de que no sea fácil para un usuario ejecutar una orden que pueda afectar todas sus copias. Vale destacar que las preocupaciones en relación a la seguridad de la información no sólo son válidas para aquellos que trabajan con contenido sensible. Conozco una biblioteca de gran envergadura que se dio cuenta de que algunos archivos maestros digitales de un proyecto habían sido redimensionados para una presentación de PowerPoint, en su ubicación de almacenamiento permanente. Los permisos de escritura que dan muchas libertades son una invitación a la pérdida de datos. Lo deseable es que usted limite la posibilidad de que alguien edite copias en el lugar en donde las está almacenando.

Ejemplos concretos

Manteniendo el espíritu pragmático y práctico de este libro, quisiera revisar algunos ejemplos sobre cómo organizaciones de distintos tamaños y que manejan diferentes escalas de contenido enfrentan los principales desafíos en el diseño de sistemas de la preservación digital para administrar la preservación de bits. Comenzaré hablando sobre lo que puede hacer un solo individuo que trabaja en una organización pequeña. Después explicaré cómo alguien en una organización mediana puede dedicar tiempo y energía a la resolución de estos problemas de manera constante. A partir de ahí, me sumergiré en el rango de complejidades que emergen en las organizaciones más grandes que



administran colecciones considerables y probablemente de carácter heterogéneo. En conjunto, estos ejemplos ilustran cómo surge el enfoque adecuado para cada organización de acuerdo con sus recursos, necesidades y objetivos. Es importante recalcar que esto puede tener como resultado enfoques muy diferentes en distintos tipos de organizaciones.

El organizador solitario en una institución pequeña

La mayor parte de las bibliotecas, archivos y museos tienen tanto un personal reducido como recursos muy limitados. Por ejemplo, una biblioteca rural promedio en los Estados Unidos tiene 1.5 empleados.⁷⁸ Muchas de estas instituciones poseen colecciones digitales de tamaño modesto, pero de carácter invaluable. Por lo que he podido recabar a través de mis conversaciones con el personal de estos lugares, es usual que las colecciones incluyan grabaciones nativas digitales de audio y video de historias orales y eventos, materiales digitalizados, como fotografías históricas o documentos, así como registros institucionales nativos digitales que continúan en crecimiento (actas de reuniones de directorio, registros de auditorías, materiales financieros). A través mi trabajo y el de mis estudiantes en la exploración del estado de estos materiales digitales en organizaciones pequeñas, hemos notado que, en gran medida, este contenido enfrenta riesgos significativos y urgentes respecto a la preservación de bits. Por fortuna, el volumen de estos materiales casi siempre permite que sea muy simple y rápido progresar.

En primer lugar, es necesario hacer un inventario del contenido digital que tiene la organización y establecer un proceso y una política para su realización y actualización. Esto puede ser tan simple como hacer una actualización mensual o trimestral de los materiales existentes. Su objetivo debe ser contar con un inventario completo de lo que tiene, al menos el número de archivos que hay en cada directorio y sus tamaños. De esa forma, es posible hacer una revisión rápida más adelante para asegurarse de que todos están presentes. Una vez hecho ese inventario, usted podrá crear copias de todo lo que necesita administrar y podrá ponerlas de forma organizada en un solo lugar para generar información sobre la integridad para todos los archivos, y entonces comenzar el mantenimiento de copias múltiples.

78 Deanne W. Swan, Justin Grimes y Timothy Owens, *The State of Small and Rural Libraries in the United States*, Research Brief 5 (Washington DC: Institute of Museum and Library Services, 2013), 7, <https://www.ims.gov/sites/default/files/publications/documents/brief201305.pdf>.



Es perfectamente posible para una organización pequeña como ésta tener todo su contenido guardado en un solo disco duro externo de tipo USB, debido al tamaño de la colección. Entonces, usted puede guardar una copia en uno de estos dispositivos, desconectarlo y ponerlo en un lugar seguro en donde no existan muchas variaciones de temperatura. Luego, puede obtener una cuenta en un servicio como Amazon Glacier, o incluso algo más simple como Dropbox. Estos servicios realizarán verificaciones regulares de su contenido y es probable que se encuentren almacenados en lugares que tienen distintos tipos de amenazas naturales que donde se encuentra usted. Para cualquier material que pueda compartir sin restricciones, usted puede subir una copia a Wikimedia Commons o al Internet Archive; ambos ofrecen alojamiento gratuito y tienen infraestructuras más bien robustas en relación con la preservación del contenido digital. Estas instituciones sin fines de lucro ofrecen alternativas tanto para mantener el acceso al contenido a largo plazo como para permitir que el contenido sea usado y descubierto.

Si una organización de poco tamaño toma estos pequeños pasos reducirá los riesgos de perder su contenido digital de forma dramática. He hablado con una variedad de personas que tienen colecciones únicas e increíbles de historias orales que viven en el disco duro de la computadora personal de una sola persona. Esas colecciones están a un derrame de café de perderse por completo. Habrá muchas más posibilidades de que se preserven en el futuro si se generan copias y se revisa el contenido de forma regular.

Una organización mediana con una colección modesta

Mientras más grande sea la organización (con más probabilidad) tendrá que conseguir más recursos para abordar los desafíos de la preservación digital. En este grupo usted puede encontrar bibliotecas de universidades pequeñas, bibliotecas públicas medianas o de sistemas de bibliotecas públicas y un rango amplio de museos de distinto tipo. Es bastante seguro que las organizaciones de este tamaño produzcan y recopilen contenido a un nivel para el que sería muy complejo, además de que consumiría mucho tiempo, llevar a cabo manualmente la copia y revisión del número de archivos, como se ha descrito para las organizaciones más pequeñas. Aunado a eso, mientras más archivos se tengan es más probable encontrar una variedad de errores al momento de copiar la información. Por esto, es una buena idea establecer un sistema que sea un poco más automatizado y sofisticado.

Esto nos lleva a considerar un par de líneas de trabajo primarias, basadas en cómo usted asigna sus recursos y cómo pretende continuar la construcción



de su organización hacia el futuro. Hasta este punto, las principales opciones son la adquisición de una herramienta que incluya alojamiento o la contratación de personal para establecer y mantener su propia estructura de almacenamiento. En muchos casos tiene sentido hacer una combinación de ambas. Su decisión en este ámbito debe estar fundamentada tanto en la forma en que está organizada su institución como en la trayectoria que estima en cuanto al crecimiento de sus colecciones. Voy a explicar mi visión sobre ambos, pero quisiera recalcar que, incluso si se trata de una organización pequeña, la preservación digital debe incluirse como un área de actividad dentro de la descripción del trabajo en el que se debe enfocar una persona. La preservación es una función principal de las instituciones del patrimonio cultural y, como tal, es esencial que esté cubierta por recursos y personal. Incluso si sólo cuenta con tres o cinco miembros de personal, alguien en ese equipo debería dedicar al menos 10 o 20% de su tiempo en asegurarse de que exista acceso al contenido a largo plazo. Es más, también se debe dedicar algo de ese tiempo a la exploración de distintas opciones y a la planificación de distintos escenarios. La idea central de la preservación digital como oficio es la noción de que requiere ser practicada por un profesional.

La primera opción es adquirir una herramienta con alojamiento. Conozco una serie de instituciones de tamaño mediano que pagan por productos como Preservica, DSpaceDirect o ArchivesDirect.⁷⁹ Ésta no es una lista exhaustiva ni una promoción de esas herramientas, sólo es ilustrativa. Otros sistemas parecidos también son útiles, pues ofrecen interfaces de usuario para automatizar gran parte del trabajo requerido en la revisión y el mantenimiento de copias de su contenido. Siempre es una buena idea adoptar el hábito de exportar de forma regular sus datos y mantener una copia local en algún lugar. De esta forma, usted sabrá que en verdad puede exportar los datos y que está protegido si algo va mal con el proveedor con el que trabaja. Los precios de varias herramientas con alojamiento están ligados de manera directa a la cantidad de contenido que se tiene, por lo que los candidatos ideales para estas soluciones tienden a ser casos en los que se almacenan cantidades modestas de datos. Sin embargo, a estas compañías, y a aquellas sin fines de lucro, también les interesa negociar contratos con clientes más grandes. Estos productos son en particular atractivos si su organización tiene una estructura en la que es más fácil contratar servicios externos que hacer lugar para un miembro permanente

⁷⁹ Para más información sobre cada uno de estos servicios, ver <http://www.archivesdirect.org/>, <http://www.dspace-direct.org/> y <http://preservica.com/>.



del personal para construir, mantener y administrar los sistemas. Una de las grandes preocupaciones en relación con estas soluciones es quedar potencialmente amarrado a un solo proveedor. Lo anterior, en especial, es importante dado que lo que usted requiere es un sistema que permita la administración de su contenido a largo plazo. Si adopta uno de estos sistemas, es necesario que establezca una estrategia de salida. También vale la pena destacar que de todos modos necesita dedicar tiempo de su personal para el uso de estos sistemas. La propuesta de valor de estos productos es que proveen automatización, mantenimiento y administración de los sistemas que ya se utilizan. Por esta razón, no es necesario que asigne tiempo de su personal para esas funciones en particular. Sin embargo, es vital recalcar que aún necesitará destinar personal adecuado para hacer uso de estos sistemas. Es importante recordar que las herramientas no llevan a cabo la preservación digital. Las personas utilizan recursos en un contexto institucional para realizar la preservación digital.

En lugar de comprar un sistema de preservación digital para realizar este trabajo, muchas organizaciones invierten en la contratación de personal dedicado a adquirir y configurar *hardware* y ejecutar *software* de código abierto para hacer su preservación digital. Dado que la preservación y el cuidado de las colecciones es una función central de una biblioteca, museo o archivo, tiene mucho sentido dar lugar a especialistas internos para que realicen este tipo de trabajo. Es difícil imaginar que muchas de estas instituciones contraten servicios externos para administrar el espacio físico en donde se encuentran almacenadas sus colecciones y que hagan el trabajo de administrar y dar acceso a ellas. Sin embargo, esto es lo que ocurre con el contenido digital de muchas organizaciones. Al momento de publicación de este libro, existen varias herramientas de código abierto que pueden realizar gran parte del trabajo fundamental de la preservación digital.⁸⁰ Por ello, mucho de este esfuerzo se trata de configurarlas y escribir guiones de comandos para conectar las piezas. Las herramientas de código abierto son gratuitas, pero son como adoptar un cachorro, no vienen libres de responsabilidades. La adopción de herramientas de código abierto, tal como la de una mascota, requiere de tiempo para cuidar y alimentar de ella. Es necesario darle una cantidad razonable de amor y atención. Al fin de cuentas, la sustentabilidad definitiva de las herramientas

80 Owens, "Open Source Software and Digital Preservation: An Interview with Bram van Der Werf of the Open Planets Foundation", *The Signal* (blog), 4 de abril de 2012, <http://blogs.loc.gov/thesignal/2012/04/open-source-software-and-digital-preservation-an-interview-with-bram-van-der-werf-of-the-open-planets-foundation/>.



y de los servicios de código abierto está anclada a los recursos y la disponibilidad del personal. Entonces, si su organización está pensando en adoptar tecnologías de código abierto para administrar una parte considerable de su trabajo, es importante entender cómo se financia y sostiene ese tipo de *software*. En muchos casos, es buena idea involucrarse en la comunidad de código abierto que rodea su proyecto o buscar formas de participar en los sistemas de gobernanza de otro proyecto para ayudar a que continúe prosperando y que las direcciones que tome en el futuro en relación con su desarrollo vayan en un sentido que apoye las necesidades y las metas de su propia organización. Junto con eso, usted también debe pensar en el tipo de *hardware* que utilizará (por ejemplo, servidores y medios de almacenamiento) en la planificación del monitoreo y la administración de dichos sistemas y en el reemplazo de sus elementos con el paso del tiempo. Si se trata de una organización relativamente pequeña, un bibliotecario o archivista con interés en esta área, un presupuesto modesto para equipo y formación continua y algunas habilidades básicas para la escritura de *scripts* de comandos pueden hacer despegar un programa como éste. Dependiendo del contexto, este enfoque puede llegar a costar más que una solución de alojamiento; sin embargo, puede traer flexibilidad y control adicionales en la administración de su contenido digital si se construye como una forma de establecer competencias dentro de la organización para apoyar este tipo de trabajo. Si usted administra una cantidad significativa de contenido, es posible que establecer sus propios sistemas y procesos sea menos costoso a largo plazo, en caso de que las colecciones continúen creciendo. Es muy probable que con el tiempo usted necesite realizar cada vez más esfuerzos para fortalecer estas competencias, por lo que hacerlo desde adentro es una buena idea.

Además de estas dos opciones, mencionaré de manera breve una tercera, que es unirse a una cooperativa de preservación distribuida. Han surgido varias organizaciones, como MetaArchive, que proveen ciertos niveles de servicios a instituciones afiliadas; sin embargo, asumen que éstas desarrollarán sus propias competencias fundamentales y su infraestructura para la mayor parte de la labor de preservación digital.⁸¹ Este tipo de iniciativas apoyan la subcontratación de una porción del trabajo a sistemas de preservación digital a través de la agrupación de recursos de las organizaciones miembro, para así aportar herramientas y servicios que de otro modo no podrían mantener.

81 Katherine Skinner y Martin Halbert, “The MetaArchive Cooperative: A Collaborative Approach to Distributed Digital Preservation”, *Library Trends* 57, núm. 3 (2009): 371-392.



Una institución grande con colecciones digitales considerables

Mientras más grande sea la organización, más complejos serán los desafíos de la preservación digital. En este caso, también son viables las mismas tres opciones mencionadas para organizaciones medianas. Sin embargo, a medida que aumenta la escala de complejidad, tanto de la estructura de la organización como del crecimiento de sus colecciones, surgen nuevos desafíos de preservación digital. Esto tiene como resultado que varias organizaciones de gran tamaño usen distintos enfoques para diferentes tipos de contenido o para sus distintas secciones. Existen dos desafíos claros de crecimiento. El primero está relacionado con las políticas organizacionales y el segundo con el aumento de los desafíos relacionados con los límites técnicos de los servicios y las herramientas. Describiré brevemente cada conjunto de situaciones.

Mientras más grande sea la organización, usted necesitará más tiempo y energía para dedicar a la planificación y la creación de políticas. También tendrá más tipos de usuarios a quienes apoyar, por lo que se volverá cada vez más importante entender sus necesidades particulares. Al mismo tiempo, es probable que su institución exista en un espacio donde hay más personas interesadas o involucradas. Debe tomar en cuenta a quienes toman las decisiones, como los decanos y los consejos, y también debe explicarles esas necesidades. A menudo hay unidades organizacionales relacionadas dentro de la burocracia: el departamento de tecnologías de la información, la oficina de comunicaciones u otras unidades que recopilan y administran el contenido o que juegan un papel decisivo en relación con las políticas del sistema de tecnologías de la información. Mientras más complejas sean estas relaciones, va a tomar más tiempo y energía trabajar con todas esas unidades para establecer una estrategia coherente. Es muy probable que los distintos grupos tengan distintos requerimientos y usted deba pasar un tiempo considerable asegurándose de que las partes interesadas se escuchen y comprendan entre sí. Por ejemplo, la palabra “archivo” significa algo muy distinto para alguien que trabaja en tecnologías de la información en comparación a lo que entiende un archivista profesional.⁸² En muchos casos, esa complejidad es tan grande que el objetivo de cohesión se vuelve demasiado ambicioso. Como mínimo, usted debe ser capaz de solucionar estos obstáculos, para que la organización funcione y trabaje lo mejor posible. También debe alinear los recursos para establecer sistemas que funcionen. En resumen,

82 Owens, “What Do You Mean by Archive? Genres of Usage for Digital Preservers”, *The Signal* (blog), 27 de febrero de 2014, <http://blogs.loc.gov/thesignal/2014/02/what-do-you-mean-by-archive-genres-of-usage-for-digital-preservers/>.



mientras más grande sea la organización, se vuelve más crítico contar con personal de tiempo completo que cuente con la autoridad y los recursos necesarios para poner en marcha un programa de preservación digital.

A medida que las colecciones aumentan de tamaño, también lo hacen los desafíos técnicos relacionados con su administración. Cuando usted trabaja con sistemas computacionales al nivel del consumidor, todo esto es hasta cierto punto sencillo. Si usted desea almacenar y gestionar datos en el rango de los *gigabytes* o un par de *terabytes* (que suele ser el caso de las instituciones que almacenan sobre todo texto e imagen), entonces el costo es lo suficientemente bajo como para no tener la necesidad de saber mucho sobre las ventajas y desventajas que pueden existir en relación con las distintas opciones de diseño de sistemas. Sin embargo, si le toca lidiar con decenas o centenas de *terabytes* o *petabytes*, es probable que necesite repensar algunas de las suposiciones que entran en juego al trabajar con sistemas computacionales. Vale la pena destacar que para las organizaciones que coleccionan contenido audiovisual digital es muy fácil alcanzar estos grandes volúmenes de datos.

Para administrar contenido de este tamaño, suele ser útil mantener sistemas de almacenamiento jerárquico. Los sistemas de cintas magnéticas tienen velocidades de uso mucho más lentas a las que estamos acostumbrados con los sistemas de discos giratorios o las memorias de tipo flash, pero siguen siendo más económicos. Entonces, la cinta magnética es una tecnología valiosa para la preservación de datos a gran escala. Como dije antes, no tiene sentido verificar la integridad de forma regular, pues esto puede maltratar y desgastar los sistemas de cinta magnética. Además de los problemas de escala que se generan en términos del volumen de almacenamiento para la preservación, vale recalcar que también existen desafíos de acceso y uso que surgen cuando hay grandes cantidades de contenido. En otras palabras, se vuelve más complejo indexar el contenido o ejecutar verificaciones de datos, simplemente por la velocidad a la que debe ser realizado el proceso. Mientras las necesidades de su institución se alejan más del mundo de los sistemas a nivel del consumidor, más necesitará entender las distintas cajas negras que componen las plataformas computacionales que se utilizan en la preservación digital. Hay mucho más que decir sobre este tema; sin embargo, los sistemas de almacenamiento y arquitectura a nivel empresarial representan un área de especialización por sí mismos, y están fuera del alcance de este libro. Una organización que necesite este tipo de trabajo debería invertir en personal de tiempo completo para administrar y mantener estos sistemas o, en su lugar, en personal que sepa lo suficiente como para manejar su contratación externa.



La autenticidad de la reproducción del contenido digital

Hasta ahora, este capítulo se ha enfocado por completo en la preservación de bits, es decir, en el trabajo necesario para asegurarse de que las copias auténticas del contenido estén disponibles en el futuro. El trabajo de gestionar copias para la preservación de bits es un problema resuelto. La mayor cantidad de esfuerzo requerido en la preservación digital está en un grupo de desafíos un poco menos definidos. Por fortuna, los retos más generales de la preservación digital son mucho menos urgentes que la preservación de bits. Si usted no realiza la preservación de bits de forma correcta hoy, ni siquiera alcanzará a entender los desafíos a largo plazo de la preservación digital, ni tendrá que preocuparse por ellos.

El desafío principal en el área más amplia de la preservación digital es que el *software* “ejecuta”. La naturaleza activa y de desempeño de ese “ejecutar” sólo es posible a través de un modelo de regresión de dependencias de diferentes piezas de *software* que, en general, están fuertemente unidas a otras piezas específicas de *hardware*. A su vez, es importante detenerse a pensar si existe suficiente contexto sobre los objetos digitales para que alguien en el futuro pueda comprenderlos (dedicaré los dos capítulos siguientes a este tema). Existen dos estrategias básicas para enfrentar el problema de los formatos: emulación y normalización o migración. Discutiré ambas de forma breve y explicaré por qué muchas organizaciones han decidido cubrir sus opciones con la normalización, pero también mantienen su material en el formato original.

Este trabajo se relaciona directamente a otros asuntos en cuanto al acceso al contenido que cubriré más adelante. Sin embargo, salen aquí a colación, ya que la definición de algunos enfoques para asegurar el acceso a largo plazo también juega un papel importante en la forma en que se diseña y administra un sistema de almacenamiento para preservación digital. De nuevo, teniendo en cuenta la naturaleza interactiva de la planeación para la preservación digital, recomiendo leer el resto de este capítulo y el que trata sobre acceso de forma paralela.

¿Los formatos se vuelven obsoletos?

Una de las preocupaciones centrales detrás del desarrollo de la teoría y la práctica de la preservación digital es que los formatos digitales que están en uso hoy en día tienen el potencial de volverse obsoletos e ininteligibles en el futuro. Incluso algo que parece muy simple, como un PDF, está sujeto a un grupo considerable de dependencias de *software* para ser reproducido de forma correcta. Las amenazas



de la preservación que pueden tener los formatos de archivo son un desafío, ya que implican tratar de predecir el futuro. A diferencia de las problemáticas alrededor de la preservación de bits, la futura usabilidad de los formatos de archivo está atada a las trayectorias venideras de la computación en la sociedad.

Los formatos son especificaciones. No son verdades fundamentales sobre los archivos. Piense en ellos como algo parecido a lo que un sacerdote lee de un texto en una ceremonia: los formatos son el canon. En contraste, los archivos varían, así como todas las distintas formas en las que las personas viven sus creencias en la vida cotidiana una vez fuera del lugar de culto. Entonces, el grado en el que un archivo en particular se ajusta o no a las especificaciones del formato siempre será un problema en sí mismo. Muchas aplicaciones de *software* que usamos a diario crean archivos que utilizamos en otras aplicaciones de *software* que son inválidos en distintos grados, de acuerdo con sus propias especificaciones. Muchos archivos son auténticamente inválidos, lo que puede ser significativo en sí mismo. Todas las conversaciones sobre formatos deben comenzar en el entendido de que éstos son convenciones y no verdades fundamentales sobre cómo deben estar estructurados los archivos.

Se ha identificado una serie de factores para evaluar la sustentabilidad de un formato digital determinado.

- **Divulgación:** si las especificaciones del formato están disponibles, entonces es posible validar los archivos y desarrollar nuevas formas de reproducirlos.
- **Adopción:** mientras más común sea el uso de un formato, será más probable que exista una variedad de implementaciones para su uso en el futuro y una demanda continua del mercado para el acceso a su contenido.
- **Transparencia:** mientras más transparente sea la estructura de la información dentro de un archivo, será más fácil su reconstrucción en caso de que ya no se pueda reproducir.
- **Documentación:** si un formato incrusta metadatos en los archivos, es más probable que alguien pueda comprenderlo, incluso si se ha perdido la habilidad de reproducirlo.
- **Dependencias externas:** mientras un formato sea más dependiente de piezas de *hardware*, sistemas operativos o *software* para poder ser reproducido, tiene mayor susceptibilidad de perder esa capacidad en caso de que alguna de ellas falle.
- **Patentes:** muchos formatos de archivo están patentados, lo que puede limitar su número de implementaciones de uso y hacer que su vida útil descanse en la compañía dueña de sus patentes.



- ***Mecanismos técnicos de protección:*** la criptografía y el *software* para la administración de derechos digitales hacen más difícil el acceso al contenido, por lo que cada uno de ellos se transforma en una dependencia externa adicional necesaria para tener acceso al contenido.⁸³

Vale la pena considerar todos estos factores al comenzar con la planificación de la preservación digital. Sin embargo, cada vez es más evidente que la adopción es el controlador principal para la accesibilidad continua. En pocas palabras, asumiendo que sobreviviremos al cambio climático antropogénico y que continuaremos teniendo industrias computacionales prósperas (admito el gran tamaño de estas suposiciones), es muy probable que podamos seguir reproduciendo cualquier formato de uso generalizado. Así, si usted cuenta con un PDF, MP3, JPEG o GIF, tendría toda la razón para creer que la gente podrá abrir esos archivos. Si esos formatos se vuelven riesgosos, su extenso uso hace que sea bastante factible que existan herramientas y servicios creados para ayudar a migrarlos cuando llegemos a ese punto.

Con esto en mente, es buena idea desempacar las prácticas y las técnicas que están disponibles, tanto para formatos menos conocidos como para quienes quieran mitigar aún más los riesgos de pérdida de la reproducción fiel del contenido digital. Las organizaciones del patrimonio cultural han desarrollado una variedad de técnicas para manejar las amenazas de la obsolescencia de los formatos. Es posible transformar un objeto en otro formato (normalizándolo o migrándolo), documentarlo a través de algún método de grabación, limitar los tipos de formatos aceptados, ofrecer distintos niveles de garantía basados en diferentes tipos de formatos o enfocarse en la emulación o virtualización de los sistemas que se requieren para proveer el acceso. En muchos casos lo más conveniente es apostar por varias de estas opciones.

La normalización y la migración

La forma más directa de evitar problemas con un formato es a través de la migración o normalización del contenido para administrar menos formatos digitales. Así, ya no se necesita lidiar con ese formato en particular. La normalización implica el cambio de formatos como parte del proceso de ingesta de

83 Para una descripción más detallada, ver Caroline Arms y Carl Fleischhauer, “Digital Formats: Factors for Sustainability, Functionality, and Quality” (ponencia, Society for Imaging Science and Technology Archiving Conference, 2005), https://memory.loc.gov/ammem/tech_docs/digform/Formats_IST05_paper.pdf.



los materiales digitales que usted ha adquirido, esto es, la transformación de un conjunto de archivos con una variedad de formatos y especificaciones distintos a un grupo definido de formatos en los que usted tiene más confianza sobre su supervivencia a largo plazo. La migración es en principio la misma idea, pero llevada a cabo en masa, cuando se decide transferir todo lo que se encuentra en un formato a otro más nuevo y sustentable.

Por ejemplo, una organización podría decidir que, si recibe archivos comprimidos en JPEG, los normalizará a un formato no comprimido como TIFF, para que no dependan de la tecnología de compresión JPEG. De forma similar, una institución podría realizar la misma acción, pero para todos los archivos JPEG en algún momento en el futuro. Este ejemplo podría justificarse como una apelación a la sustentabilidad; sin embargo, es muy probable que sea absolutamente innecesario. Los archivos JPEG son usados de forma masiva, por lo que es difícil imaginar un momento en que no se podrá reproducirlos. Recuerde que la preservación digital se trata de la administración de riesgos y la mitigación de amenazas. Entonces, aunque un formato pueda ser mejor en teoría, en términos de sustentabilidad de formatos, lo que resulta en verdad relevante es la probabilidad de que los riesgos se conviertan en un problema al que sea necesario responder. Existen suficientes asuntos más urgentes como para pasar el tiempo respondiendo a cuestiones imaginarias que tienen muy pocas posibilidades de ocurrir. No obstante, si en algún momento parece que será difícil abrir archivos JPEG, entonces tendría sentido pensar en migrarlos en masa a un formato más moderno. Dada la forma en que han madurado los formatos de imagen en las últimas décadas, esto no me preocupa demasiado, pues los requerimientos para la reproducción de imágenes en una gama variada de navegadores web ha permitido moderar algunos de los problemas que eran alarmantes antes de la existencia de la red.

El desafío más grande para una organización que desea normalizar o migrar su contenido a un formato nuevo es la garantía de la calidad. ¿Cómo puede usted estar seguro de que lo que importa preservar del archivo original que tiene está presente en el archivo nuevo que está creando? La única forma de tener esa certeza es dedicando mucho tiempo y energía al establecimiento de prácticas y métodos para garantizar la calidad. Cualquiera de éstos también requerirá de un sistema operacional sólido para definir con claridad los aspectos que de verdad son significativos del material con el que usted trabaja. Esto es una buena opción, ya que existen muchos aspectos de un objeto digital que pueden ser significativos; en la mayor parte de los casos, esto es un desafío. De esta manera, las organizaciones que realizan normalización o



migración suelen guardar los originales o dedicar esfuerzos y cuidados considerables al establecimiento de procesos para garantizar la calidad. Si en vez de normalizar los archivos, usted utiliza el mismo procedimiento para crear derivados en otros formatos de forma rápida, puede justificar el dedicar menos tiempo y cuidado en garantizar la calidad, y de todas formas puede estar protegido ante ciertos problemas de obsolescencia de formatos.

La documentación

Como mencioné antes, en el capítulo sobre el propósito de la preservación, hay casos en los que existen sutilezas en la reproducción de la información que se consideran importantes, por lo que uno desea registrar con exactitud cómo algo se veía en una pantalla o cómo sonaba a través de la instalación de un *hardware* en particular. Los ejemplos de *Form Art* y *World of Warcraft* del capítulo anterior ilustran el valor de la documentación: tomar capturas de pantalla o utilizar herramientas que puedan crear grabaciones de audio o video de lo que se está reproduciendo. En ese caso, usted estará creando nuevo contenido de forma funcional, por lo que tendrá muchas más opciones para decidir los formatos de audio, video o imagen que va a utilizar.

Para algo que sea dinámico o complejo en particular, o para algo que necesita ser reproducido perfectamente al nivel de los píxeles para mantener su significado, es muy probable que los formatos de audio, video e imagen sean mucho más estables y confiables que cualquier otro tipo de formato. Por supuesto, esta documentación no será la obra o el objeto original con sus propiedades interactivas, por lo que, si éstos fueran los únicos elementos adquiridos, habría una pérdida considerable de sus cualidades y características reales. Al igual que en el caso de la generación de derivados que incluye el mantenimiento de los originales, ésta puede ser una forma aceptable de reducir el riesgo. Es decir, usted puede documentar un objeto o una serie de objetos, y las imágenes y grabaciones resultantes pueden convertirse en la referencia para que futuros usuarios lo comparen con una reproducción del objeto original en algún sistema para verificar su calidad o autenticidad.

Limitar los formatos aceptados u ofrecer niveles de garantía para los formatos

Una de las mejores formas de lidiar con los posibles problemas de los formatos es tratar de anticiparse a ellos desde el inicio. Esto es algo que sólo es posible en situaciones donde los involucrados en la preservación del contenido pueden intervenir desde el principio en la cadena de producción del contenido. Por



ejemplo, si usted administra un depósito institucional, tiene la habilidad de solicitar que los documentos sean entregados en formato PDF, en vez de aceptar distintos tipos de formatos de procesador de texto. De forma similar, si usted gestiona el archivo en una organización de producción de medios (como una estación de TV o de radio, un periódico o publicación digital) y puede instaurar la entrega y validación de archivos a través de un sistema para archivar como parte del flujo de trabajo continuo, entonces puede incrustar la administración de los formatos y la garantía de la calidad en las operaciones diarias de su organización. También puede forzar la validación de los archivos al momento en que la gente sube los materiales y, si los archivos fallan, podría también hacer que sea un requisito que el usuario lo arregle y lo vuelva a subir.

Algunas instituciones, para ser menos restrictivas, ofrecen distintos niveles de garantía para diferentes tipos de archivos digitales. Por ejemplo, la Michigan University proporciona tres niveles de garantía: alto, medio y bajo, para archivos en distintas categorías de formato.⁸⁴ El valor de este enfoque es que sugiere y recomienda formatos con los que probablemente sea más fácil trabajar para los fines de la preservación, pero no rechaza ningún contenido que no alcance el nivel mínimo.

La emulación y la virtualización

La última estrategia para manejar las problemáticas relacionadas con los formatos de los archivos y sus dependencias es hacer muy poco de forma inmediata y, en su lugar, enfocarse en las tecnologías que, en el futuro, puedan virtualizar los sistemas antiguos o recrearlos a través de la emulación para reproducir el contenido que no podría hacerlo en sistemas operativos comunes. Se ha progresado de forma considerable en esta área, y varias organizaciones han comenzado a experimentar con estrategias basadas en la emulación y la virtualización. En la práctica, éste es un método de acceso que permite reproducir el contenido.⁸⁵ Sin embargo, también es una consideración fundamental para la planificación del acceso, por eso se sitúa en este capítulo, porque es una de las preocupaciones básicas de la preservación digital de los formatos.

La emulación y la virtualización son emocionantes, pues sugieren una manera de avanzar dejando a los bits en paz. Cualquier transformación de

84 “Deep Blue Preservation and Format Support”, University of Michigan Libraries, acceso el 28 de abril de 2021, <https://deepblue.lib.umich.edu/static/about/deepbluepreservation.html>.

85 Dianne Dietrich *et al.*, “How to Party Like It’s 1999: Emulation for Everyone”, *Code4Lib Journal*, núm. 32 (2016), <http://journal.code4lib.org/articles/11386>.



objetos digitales que involucre migración, normalización y documentación es un proceso que en realidad crea nuevos objetos digitales. Por ello, la posibilidad de dejarlos tal cual son y sólo trabajar en la construcción de sistemas que puedan reproducirlos es muy llamativa. Es importante decir que la emulación y la virtualización también abren la posibilidad de futuras migraciones a través del mismo método, es decir, virtualizando un sistema antiguo para reproducir el contenido y luego utilizar la reproducción de ese sistema para migrarlo a un formato nuevo.

Como gran parte del enfoque en este libro, la respuesta a muchas de las preguntas sobre la preservación digital comienzan con la frase “depende de”. En la mayoría de los casos vale la pena apostar por múltiples opciones. El costo de mantener contenido en múltiples formatos es lo suficientemente bajo como para justificar el esfuerzo invertido en mantener tanto los archivos originales como los normalizados o los derivados. Entonces, es bueno mantener formatos múltiples del contenido como una opción predeterminada y sólo considerar no seguir este método cuando se vuelve inaccesible en términos económicos. Si usted sigue este camino, es importante que establezca procedimientos y procesos para articular las relaciones entre estas formas de datos derivadas.

No hacer daño, proteger los bits, monitorear a escala

Para cerrar este capítulo, quiero sugerir que todos los involucrados en esta tarea tomemos una especie de juramento hipocrático para la preservación digital en relación con los formatos de los archivos y acordemos, de forma general, que “lo primero es no hacer daño”. Una vez que usted haya recibido el contenido que desea preservar, haga una revisión mínima de lo que tiene, luego transféralo a su sistema de almacenamiento de preservación, haga el trabajo requerido para asegurarse de que tiene múltiples copias de él y revíselas de forma regular. Esto le permitirá hacer archivos derivados en masa cuando se necesiten. Si, por el contrario, usted cree que existe una razón para hacer gran parte de este trabajo con los materiales desde el inicio y es necesario transformarlos o normalizarlos, también puede hacerlo, pero es importante que tenga una buena justificación para ello.

Las mayores pérdidas digitales a gran escala que han ocurrido y siguen ocurriendo no están relacionadas con los formatos de los archivos, sino con la preservación de los bits. Si usted pierde los bits, reescribe sobre los archivos, deja caer un disco duro o ocurre un desastre y todas las copias se encuentran en un solo lugar y son destruidas, entonces, perderá todo. Por eso, mi consejo principal para alguien que recién comienza en este campo es asegurarse de que



ha cubierto los requerimientos básicos de la preservación de los bits, incluso antes de pensar o preocuparse por los problemas de los formatos digitales.

Si usted entrara a un archivo en donde todas las cajas están en el suelo en un área de riesgo de inundación y se preocupara por los daños que las grapas o los sujetapapeles le están haciendo a los documentos, estaría haciendo una mala práctica archivista. Usted debe priorizar. Comience levantando las cajas del suelo. De forma análoga, no hay que cometer este error con la información digital. Necesitamos ser consistentes y asegurarnos de que no estamos actuando de la misma manera con los equivalentes digitales de las grapas y los sujetapapeles mientras las cajas digitales están metafóricamente en el suelo. Una vez que tenga implementado un sistema básico para mantener las copias de su contenido, será más fácil mirar toda la colección para tomar decisiones sobre acciones que se puedan aplicar de forma sistemática. Es decir, usted puede resolver que lo mejor es generar derivados de archivos de texto o archivos HTML de todos documentos de los distintos procesadores de palabras que existen en una colección. De esta forma, podría proveer de manera más fácil una búsqueda de texto completo a partir de estos archivos e incluso reproducir esos documentos de forma directa en cualquier navegador web que utilice el usuario.⁸⁶ Usted podría hacer esto en una sola ocasión y luego establecerlo como un proceso automatizado que ocurre cada vez que ingresa contenido nuevo. Este tipo de derivados tiene un valor múltiple, pues sirve tanto para garantizar el acceso a largo plazo de ciertas partes del contenido como para hacerlo más accesible y darle usos variados de forma inmediata. Si usted hace este tipo de derivados, por favor, en general, no se deshaga de los archivos originales. Guardar los originales usualmente requiere un espacio de almacenamiento trivial, y sin ellos será muy difícil saber si perdió algo al crear un archivo derivado.

86 Para consultar un ejemplo de este tipo de enfoque, ver Owens, “Mecha-Archivists”.

Capítulo 7

La organización y la descripción de los objetos digitales

Know Your Meme es una comunidad en línea que documenta los memes de Internet (chistes, imágenes y frases que se propagan en línea). El sitio contiene entradas enciclopédicas sobre memes individuales, con comentarios de los usuarios ligados a ellos, y una serie de videos de la misma temática. Todo esto se alimenta mediante una base de datos de soporte que vincula y asocia los recursos individuales. Entonces, si usted deseara preservar Know Your Meme, ¿qué debería hacer en términos de organización y descripción? Como se discutió antes, es fundamental establecer sus propias intenciones o las de su institución en cuanto al porqué de la recopilación de esos materiales, lo que dará información respecto a qué se debe integrar a su colección. Sin embargo, esto todavía deja abiertas preguntas significativas sobre cómo proceder para describirlo y organizarlo. ¿Debería tratarse como un sitio web individual con un único registro? ¿Debería hacerse un registro por cada entrada de enciclopedia? ¿Se debería sacar la serie de videos y tratarla como una publicación en serie? Este capítulo trata de desentrañar las cuestiones necesarias para responder justo este tipo de interrogantes.

Una de las mejoras clave que las instituciones culturales dan a las colecciones a su cargo es justamente la organización y la descripción. Los archivistas usan el término “organizar”. Éste puede parecer un poco extraño para los profesionales de bibliotecas y museos que, en términos generales, no han necesitado involucrarse demasiado con las estructuras internas de sus materiales. Como sea, todas las organizaciones del patrimonio cultural necesitan poder organizar lo que tienen y desarrollar métodos para describirlo de modo que sea posible utilizarlo y acceder a él. La organización y la descripción son el proceso



mediante el cual las colecciones se hacen reconocibles, inteligibles y legibles para los futuros usuarios.

Utilizo los términos archivísticos por tres razones. Primero, los archivistas han desarrollado métodos para trabajar con grandes cantidades de material. Segundo, también se ocupan de materiales que vienen con una lógica y una estructura basadas en cómo se crearon o produjeron. Tercero, las colecciones de archivos a menudo son muy heterogéneas: incluyen una variedad de documentos, objetos y formatos de medios. Como resultado de estos puntos, la práctica archivística contemporánea funciona desde una relación muy directa entre la organización del contenido y las descripciones que deben coincidir con esas estructuras organizativas. Estos puntos son cuestiones relevantes para la mayoría de los escenarios en los que las bibliotecas, los archivos y los museos trabajan con contenido digital. La práctica de archivo se fundamenta en una apreciación del “orden original”, que respeta la secuencia y la estructura de los registros tal como se utilizaron en su propio contexto. Lo anterior permite dos cosas. La primera es que mantiene relaciones contextuales entre los materiales de una colección que pueden ser útiles para comprenderlos. La segunda es que aprovecha esa estructura y ese orden inherentes como un medio para proporcionar el acceso. Es decir, ahorra tiempo y esfuerzo invertido en reorganizar el material, pues permite que el orden original del contenido persista.⁸⁷ A lo largo de este capítulo, me moveré entre la organización y la descripción como actividades interrelacionadas. En última instancia, el objetivo de todo esto es proporcionar acceso y uso, materias en las que profundizaremos en el próximo capítulo.

Con esto en mente, proporciono un mapa para el capítulo. Comienzo intentando tranquilizar a los lectores. Para empezar, lo mejor es adentrarse en el trabajo de organizar y describir el material digital con los métodos y las prácticas existentes. Después, profundizaré en el enfoque “más producto, menos proceso” (MPLP) para procesar, organizar y describir colecciones de archivo. He mencionado brevemente este planteamiento en el cuarto capítulo, pero dada su importancia y su relevancia para trabajar con el contenido digital, me centraré un poco más aquí. Después, utilizaré como ejemplo una adquisición digital nativa de las Stanford University Libraries para analizar cómo la infraestructura del MPLP puede permitirnos realizar intervenciones estratégicas mínimas para organizar el contenido y tener enfoques de muy alto nivel con el fin de describir el contenido digital.

87 Society of American Archivists, *Dictionary of Archives Terminology*, s. v. “Original Order”, acceso el 28 de abril de 2021, <https://www2.archivists.org/glossary/terms/o/original-order>.



Después me adentraré en una serie de ejemplos adicionales para aclarar cómo las características clave de la información digital discutidas antes dan lugar a desafíos y oportunidades para la organización y la descripción. Usando el ejemplo del sitio web Know Your Meme, me concentro en lo difícil que es definir los límites de los objetos digitales. Aquí, la naturaleza de las bases de datos de los objetos digitales define los enfoques necesarios para la organización y la descripción. Más adelante, a través de dos tuits de #Ferguson, demuestro hasta qué punto los objetos digitales pueden o no describirse a sí mismos. Desde estos contextos, sostengo que necesitamos pasar a un enfoque que respete una multiplicidad de órdenes, variantes y copias. Exploro cómo funciona esto mediante algunas discusiones en la investigación que giran en torno a la manera en que la gente común maneja y organiza sus fotografías digitales, así como a través de una exposición de GIF animado del Museum of the Moving Image. Como complemento, analizo los enfoques de las aplicaciones y los repositorios institucionales dedicados a la recopilación de la historia oral que transmiten su trabajo de descripción a los contribuyentes y los usuarios.

Cierro el capítulo con una discusión sobre los principales cambios que me parece que moldearán el futuro de la organización y la descripción de los materiales digitales. Cada vez más, operamos en un mundo de colecciones y descripciones relacionadas entre sí y conectadas en la red. Conforme incrementa el contenido disponible en línea, a un simple clic de distancia de los materiales de un gran número de instituciones, las colecciones institucionales individuales, de manera creciente, se han convertido en parte de una colección global vinculada. En este contexto, los materiales digitales se describirán a sí mismos y mutuamente con más frecuencia. Esto sugiere la necesidad de alejarse de algunas nociones tradicionales de las funciones de la catalogación para avanzar a un mundo en el que los bibliotecarios, los archivistas y los curadores funcionen como manipuladores de datos interconectados.

Entonces, si el resumen sobre la dirección que lleva el presente capítulo ya se siente un poco complicado, no se preocupe. Por complejas que parezcan las cosas, estamos preparados para adentrarnos y comenzar de manera simple.

Partir de una suposición de continuidad

Empiece por no olvidar lo que ya hacemos. Si tiene una foto digital, descríbala de la misma manera en que lo haría con una foto analógica. El registro del catálogo de un libro electrónico debe encontrarse junto con el registro de un libro en un estante. Un asistente para la búsqueda de archivos puede (y debe) integrar de forma directa la información sobre los componentes nacidos digitales



en una colección de documentos personales. En muchas áreas pragmáticas, la introducción de los medios digitales no debe alterar nuestras prácticas para gestionar la información.

Las fotos digitales, los libros electrónicos o el material de archivo nacido digital puede almacenarse en algún otro sistema. Usted puede tener requisitos diferentes para proporcionar acceso a esos materiales. Sin embargo, existen ventajas significativas para entrelazar las prácticas de su organización y su descripción. Vale la pena reiterar que “lo digital” no es un universo de objetos y experiencias del todo diferente. Los objetos digitales son parte de la continuidad de la recopilación de formas y formatos tradicionales y de medios, por lo que, para empezar, es importante no intentar separar artificialmente las cosas sólo porque son digitales.

Dado que gran parte del trabajo de la organización y la descripción debe integrarse en las prácticas existentes, no me voy a centrar mucho en este capítulo en esquemas y estándares para la descripción como Dublin Core, MARC, Bibframe, RDA, Encoded Archival Description y DACS. Más allá de eso, no hablaré mucho sobre las herramientas específicas que se utilizan para administrar y organizar metadatos descriptivos. Si usted ya tiene estándares y herramientas en uso, entonces tiene sentido que establezca de forma predeterminada continuidad con esas prácticas. En este contexto, la cuestión clave es incluir materiales digitales en los flujos de trabajo y los procesos existentes para esas herramientas y estándares. Lo que sigue se centra, en principio, en pensar en la división conceptual de los objetos digitales en unidades significativas y en el nivel de descripción para comprometerse dentro de cualquier esquema y sistema que se esté utilizando. Para aquellos que no están usando ninguna herramienta para administrar metadatos descriptivos, sugeriría explorarlos, pero, incluso sin ese contexto, lo que viene a continuación será útil para profundizar en las cuestiones intelectuales en las que tendrá que reflexionar, más allá de cualquier herramienta o esquema.

Con todo esto señalado, acerca del mantenimiento de la continuidad, vale la pena subrayar que las prácticas para organizar y describir las colecciones todavía están, de una manera muy real, poniéndose al día con el presente. Con especial énfasis en las prácticas de archivo, los cambios que apuntan hacia un nuevo modelo para abordar estas tareas han sido en particular relevantes en los últimos diez años.



More Product, Less Process: la revolución inconclusa

Hace más de una década, Mark Greene y Dennis Meissner convocaron a los archivistas a utilizar el modelo MPLP, ya mencionado, para la integración de la disposición, la preservación y la descripción de las colecciones de archivos.⁸⁸ Las preguntas que hicieron sobre la práctica de archivo son muy relevantes para trabajar con el contenido digital dentro de una amplia gama de instituciones del patrimonio cultural. En dicha práctica, existe un consenso generalizado en torno a la necesidad de avanzar en la dirección MPLP. Sin embargo, el movimiento todavía es bastante lento. En la realidad, este replanteamiento del trabajo de los archivos aún se está abriendo paso, pues tampoco está particularmente bien integrado al trabajo de la preservación, la organización y la descripción de los objetos digitales. Explicaré un poco el método MPLP para proporcionar el contexto a los lectores que no estén familiarizados con él. Dado que este trabajo se centró sobre todo en colecciones de archivos analógicos, también intentaré traducir sus implicaciones al trabajar con materiales digitales. El problema central que el MPLP pretende abordar es que los archivos han tardado demasiado en procesar las colecciones, y esto ha devenido en grandes retrasos de colecciones no procesadas que, en general, no están disponibles para los usuarios.

Los archivistas han sostenido durante mucho tiempo que su trabajo no debe incluir la descripción de cada elemento individual. En cambio, se centran en descripciones del contenido a niveles más altos de agregación: de colección o de serie. Dada la escala con la que trabajan, la descripción del elemento individual es ineficiente. Al mismo tiempo, muchas sugerencias de toda la vida para el arreglo y la preservación de las colecciones de los archivos han requerido que se trabaje mucho a ese nivel. El argumento central de MPLP es que las actividades de la preservación, la organización y la descripción deben ocurrir en armonía, con un nivel común de detalle. Así es que, Green y Meissner sugieren que: “el buen procesamiento se realiza con una pala, no con pinzas”.⁸⁹ El resultado de esta recomendación es que, para el trabajo de la preservación, la organización y la descripción, el objetivo debe ser “el mínimo dorado”, es decir, el mejor nivel de organización y descripción que implique la menor cantidad de trabajo para obtener registros en un formato que los haga utilizables. En la práctica, esto debería producir un nivel mínimo de descripción para la recopilación de todo lo que una institución posee y adquiere antes

⁸⁸ Greene y Meissner, “More Product, Less Process”, 208-263.

⁸⁹ *Ibid.*, 240.



de continuar. Junto con eso, Green y Meissner sugieren que las organizaciones, en general, deben evitar poner sus preocupaciones de preservación antes que las de acceso. En este contexto, para cualquier colección, una institución debe obtener una descripción a nivel de colección y producir una lista de contenedores para abrirla a los usuarios. Estos autores también señalan que “las colecciones no procesadas deben asumirse abiertas para los investigadores. Punto”.⁹⁰ Su modelo funciona como una especie de ampliación progresiva para el trabajo con colecciones. La labor que se hace para organizar y describir una colección debe ser mínimamente invasiva y no debe retrasar el manejo del contenido que los usuarios puedan empezar a hacer. En el enfoque MPLP, las instituciones deben sentirse cómodas para tomar decisiones en cuanto a la información confidencial a nivel de lote; se decide sobre un pajar, no sobre cada una de las pajas. Esto significa que aquellos que trabajan en archivos y otras entidades que aplican el MPLP deben enfocarse en perspectivas que permitan buscar información confidencial sin mirar cada elemento individual.

Si tomáramos en serio este planteamiento para trabajar con el contenido digital, tendríamos enfoques muy diferentes para manejar los materiales digitales nativos de lo que la mayoría de las organizaciones del patrimonio cultural están haciendo ahora. Usted adquiriría contenido, crearía una breve descripción a nivel de la colección, luego produciría una lista de contenedores (que como archivos y carpetas digitales tienen la capacidad instantánea de proporcionar una lista; esto es bastante trivial) y luego continuaría y haría que la colección sea lo más accesible posible. En muchos casos, usted simplemente puede subir los archivos a un directorio en la red. El punto de este enfoque es que el “mínimo dorado” debería ser el valor predeterminado y que todo lo que esté más allá de este nivel de esfuerzo debería ser justificado.

Según mis observaciones a lo largo de los años, tengo la impresión de que casi nadie está haciendo esto con los materiales digitales. Si bien no puedo estar seguro, creo que la razón por la que pasa esto es consistente con el diagnóstico de Greene y Meissner en cuanto a que los archivistas continúan enfocando demasiado tiempo y energía en el trabajo a nivel del objeto individual. En sus palabras:

Gran parte de lo que pasa por organización del trabajo de procesamiento es en realidad administración del hogar demasiado entusiasta y a lo grande. Nuestra meticulosidad profesional y la renuencia a ser percibidos como descuidados

⁹⁰ *Ibid*, 252.



o indiferentes por los usuarios y otros han alentado una fijación generalizada en ciertas tareas que ni siquiera es necesario realizar.⁹¹

Por más sofisticada que sea una colección de material digital, en el nivel más básico, es posible hacer una copia y ponerla a disposición de los usuarios. Si usted se apega al “mínimo dorado”, es probable que dedique mucho menos esfuerzo y traslade el contenido a los usuarios más rápido. Para ilustrar esto, daré un ejemplo de un archivo digital nativo que se abordó de esta manera.

El archivo 4chan de la Stanford University

El tablero de imágenes 4chan es un sitio web comunitario en línea en el que la mayoría de los usuarios publica mensajes e imágenes de forma anónima. Ha desempeñado un papel importante en el desarrollo de varias culturas y memes de Internet y también está conectado con grupos como Anonymous y otros movimientos de la derecha alternativa. Una característica clave del sitio es que borra de forma regular su contenido. Sin embargo, algunos de sus usuarios han querido preservarlo.

Si bien 4chan continúa borrándose, existe una copia de una parte importante de su contenido que ahora está disponible en el repositorio digital de la biblioteca de la Stanford University. Esta colección, de aproximadamente cuatro *gigabytes* de datos, está disponible para que cualquiera la descargue y contiene más de 25 mil hilos de discusión. El resumen señala que la colección incluye un mensaje de quien adquirió estos documentos en el que explica un poco sobre su alcance y cómo y por qué los recopiló. En la página del repositorio de Stanford se pueden ver los metadatos descriptivos sobre el formato (Manuscrito/Archivo), la extensión de tiempo (2007-2013) y el contribuyente (identificado como Okeh, Ndubuisi [coleccionista]).

En este caso, el encargado de la recopilación buscaba un hogar para esta colección, y dada la plataforma del repositorio digital existente en Stanford y su confianza para hacer accesible su contenido, fue fácil adquirirla y registrarla en el nivel de colección. Stanford actúa como el repositorio y hace que los datos estén disponibles para que otros los descarguen, exploren, manipulen y hagan lo que quieran. Su adquisición no requirió mucho en términos de recursos y es probable que en el futuro tampoco necesite tanto para mantenerse en un nivel básico.

91 *Ibid.*, 241.



Esta es una manera muy clara y directa de realizar la organización y la descripción de algo. En primer lugar, usted puede comenzar por no organizar nada. El orden original de la colección de 4chan se ha mantenido y, como se trata sobre todo de texto digital, permite hacer búsquedas. En segundo lugar, usted debe crear un registro al nivel de colección y proporcionar la apertura de acceso que pueda ofrecer en términos legales y éticos. Si en algún momento en el futuro usted desea regresar y seccionar esto en una estructura de un nivel inferior para algún uso específico, podrá hacerlo. Sin embargo, este enfoque sirve para ofrecer lo más mínimo del “mínimo dorado” al tiempo que le permite proporcionar el acceso al contenido. Esto puede servir como un ejemplo de lo indispensable que se necesita para comenzar.

El límite de los objetos digitales: ¿es un sitio web un elemento?

El ejemplo del archivo 4chan es útil para ilustrar qué tan pequeña puede ser una intervención. Primero usted debe comprender lo que está recibiendo y debe manipularlo lo menos posible. Después debe crear y agregar un registro. Luego debe seguir adelante. Pero alguien podría cuestionar, con justa razón, si este modelo ofrece lo suficiente para alcanzar ese mínimo. ¿Cómo se puede tratar algo llamado archivo como un solo elemento? En particular, los archivistas han desarrollado una amplia gama de prácticas en torno a las diferencias entre los niveles de descripción asociados con el nivel de recopilación, los niveles intermedios variables y el nivel del elemento. Como se discutió en el capítulo dos, la toma de decisiones sobre los límites de los objetos digitales es sorprendentemente complicada. Para ilustrar este punto, hablaré sobre un objeto que forma parte del Digital Culture Web Archive en el American Folklife Center.

En 2014, el American Folklife Center de la Library of Congress comenzó una iniciativa para recopilar y preservar los registros de las culturas de Internet. Tuve la fortuna de poder ayudar a definir y desarrollar el alcance de esta colección.⁹² Al igual que muchas iniciativas de otros archivos web, la Library of Congress trabaja con el Internet Archive para rastrear sitios web y luego describir sus colecciones en línea con base en la idea de que cada uno forma parte de una colección. En este caso, cuando el Digital Culture Web Archive se ponga a disposición del público, será como una colección en línea compuesta por sitios web individuales que se han archivado de manera continua a

92 Owens, “Creepypastas, Memes, Lolspeak & Boards: The Scope of a Digital Culture Web Archive”, *Folklife Today* (blog), 26 de septiembre de 2014, <https://blogs.loc.gov/folklife/2014/09/scoping-a-digital-culture-web-archive/>.



través del tiempo. Cada sitio web se presenta conceptualmente como un solo elemento. Esto parece sencillo, pero al analizar Know Your Meme, uno de los sitios de esta colección, quedará claro lo complicado y multifacético que puede ser cualquier sitio web si se toma como un objeto.

Know Your Meme (Conoce tu meme)

Ahora ya contamos con algunas de las herramientas contextuales y conceptuales necesarias para volver al ejemplo de Know Your Meme, discutido en la introducción de este capítulo. Si bien en verdad es realmente un sitio web que se puede recopilar, preservar y representar como tal, también está compuesto por una variedad de diferentes tipos de objetos adicionales que bien podrían organizarse y describirse de acuerdo con diferentes tradiciones y enfoques.

Como cualquier wiki, éste funciona como una enciclopedia generada por los usuarios: cada página tiene una serie de autores e incluye un conjunto de enlaces y citas. De igual forma, como cada una de las entradas individuales en la wiki, cada meme tiene un título, autores y una gama de metadatos. El hecho de que sea una enciclopedia apoya la noción de que se trata de un solo objeto que podría catalogarse como alguna vez se hizo con el conjunto de tomos de la *Enciclopedia Británica* encuadernada. Sin embargo, eso no es todo en lo que respecta al sitio de Know Your Meme. Junto a esto, los videos están estructurados en episodios y temporadas, lo que sugiere la posibilidad de organizar y describir cada uno de ellos como parte de una publicación en serie. Más allá de eso, todos los elementos del sitio, es decir, las entradas, cada episodio de las series de videos, cada imagen, pueden tener comentarios de los usuarios con su propia fecha de publicación y autor. En un sentido funcional, todos estos diferentes recursos existen como micropublicaciones en la base de datos subyacente que alimenta el sitio. Uno podría imaginar que es posible tratar cada uno de los elementos, como los episodios, los comentarios, las entradas, las imágenes, etc., de manera individual, pues actúan como parte de la colección de cosas interrelacionadas que es Know Your Meme. Además, así como desempaquetamos todas las partes del sitio web de la banda Marzapan en el CD en el segundo capítulo, uno podría desempaquetar cualquiera de las entradas del sitio, las cuales contienen un rango de archivos individuales, que a su vez tienen su propio nivel de metadatos.

Como ya mencioné, dentro del Digital Culture Web Archive, Know Your Meme es tratado como un elemento particular. Sin embargo, esto deja fuera otra dimensión del sitio. Mediante la tecnología usada para archivar contenido en la red, ese elemento es recopilado de forma regular. Así, la Library of



Congress tiene copias del sitio tal como apareció para el rastreador web en fechas y horas específicas. Esto es significativo para el sitio Know Your Meme, ya que la página de inicio presenta memes que están en tendencia en un momento específico, lo que representa un aspecto diferente, y potencialmente significativo, del sitio: las páginas individuales en el sitio cambian y se desarrollan con el tiempo, por lo que archivarlas garantiza que esos cambios estén (hasta cierto punto) documentados y capturados. Por lo tanto, no sólo cada pieza del sitio puede ser considerada de forma individual, sino que también cada uno de esos elementos que se documentan en el tiempo a medida que van cambiando. Toda esta complejidad se oculta en gran medida con las prácticas y métodos de la labor de archivar contenido web. Esto es conveniente ya que la mayor parte de esas relaciones están a su vez contenidas o hechas visibles en los metadatos de los expedientes archivados y en la estructura de la URL del sitio, a medida que se conserva (cada entrada de meme está en */meme* y cada video en */videos*). Entonces, sería posible extraer estos subobjetos en una fecha futura si alguien quisiera representarlos o estudiarlos de esa manera. Es decir, cualquier usuario futuro podría sacar del sitio web archivado la secuencia de videos y las fechas en que fueron publicados para presentarlos como una colección específica de videos en serie. Dicho esto, usar la tecnología para recopilar sitios web mediante rastreadores web implica comenzar desde una URL inicial y luego saltar desde allí hacia los enlaces que se encuentran allí y en las subsecuentes. Por lo tanto, los sitios web archivados que se derivan de este proceso no son una copia completa de todo lo que está dentro de la base de datos de Know Your Meme y del servidor web en un momento dado. Sin embargo, si se deseara ese tipo de objeto digital, se podría haber abordado la adquisición del sitio no como tal, sino como una base de datos.

Entonces, ¿Know Your Meme es un elemento? Sí. También es una colección, una publicación multimedia en serie, una base de datos y un conjunto de archivos. En cualquier caso, tratarlo como un elemento particular es una forma eficiente de adquirirlo y organizarlo y es una forma relativamente mínima de describirlo. Al utilizar Way Back Machine (el *software* creado para interpretar sitios web archivados) es posible que los usuarios naveguen por el sitio tal como apareció para el rastreador web en un día y hora determinados. De acuerdo con la tecnología desarrollada para archivar contenidos web, los archivos reales que componen el sitio web están organizados de manera algo aleatoria en Web ARChive (WARC). Éstos son directorios que contienen todos los archivos individuales capturados mientras se rastrea un sitio en el orden en que se descargaron. Hay casos en los que podría tener sentido dedicar tiempo y



energía a recopilar, organizar y describir cada temporada individual de la serie de videos. Sin embargo, el archivo web ofrece una forma bastante rápida y directa de capturar la naturaleza multifacética del sitio y hacerlo reservable y accesible. Además, muchos de esos subconjuntos de elementos dentro del sitio podrían extraerse y ponerse a disposición de los usuarios como colecciones derivadas, en algún momento en el futuro, punto al que volveremos en el próximo capítulo sobre acceso.

El caso de Know Your Meme es muy similar a una variedad de otras situaciones, como la *laptop* de Rushdie y los disquetes de Larson y de Sagan: cada objeto es al mismo tiempo un elemento y una colección de elementos. En cierto nivel, esta clase de objetos compuestos son ideales para los que usamos MPLP. Es posible hacer y replicar copias perfectas de ellos como agregados y luego dejar que la descripción del elemento de nivel inferior emerja del orden y la estructura de los subobjetos/elementos. Si se vuelve útil o necesario en el futuro tener un conjunto de subcomponentes extraídos y descritos con más detalle para una comunidad de usuarios en particular, entonces será fácil hacerlo. Esto se debe en parte a que los objetos digitales tienen una naturaleza autodescriptiva. Abordar algunos ejemplos de Twitter nos ayudará a ilustrar aún más esta cualidad y los matices adicionales necesarios para establecer los límites de los objetos digitales.

Los objetos digitales semiautodescriptivos

La historia, según me informó un colega, es que poco después de que la Library of Congress firmara un acuerdo con Twitter para comenzar a archivar tuits, un catalogador preguntó: “¿Quién catalogará todos esos tuits?”. Aunque es sólo una anécdota, sirve de ejemplo del tipo de folclor que se recibe en una gran institución y como muestra de las diversas dificultades que enfrentan los catalogadores de todo el mundo. Sea verdadera o falsa, la historia expresa un punto de vista útil para el cambio de perspectiva que las instituciones del patrimonio cultural deberían estar generando. Las grandes organizaciones tienen millones o cientos de millones de artículos físicos: son muchas cosas. Sin embargo, ese contexto hace que la idea de trabajar con decenas de miles de millones de objetos (como Twitter) sea en verdad desalentadora.

Como la mayoría de los objetos digitales, los tuits vienen con una gran cantidad de metadatos transaccionales: marcas de tiempo, nombres de usuario, identificadores únicos, enlaces a URL. Asimismo, con algunas notas del alcance y el contexto generales sobre una colección dada de tuits, éstos pueden, en buena medida, describirse a sí mismos. De manera ilustrativa, vale la pena mirar y



desempacar un ejemplo de este principio, ya que se relaciona con otros objetos digitales. Para nuestros propósitos, voy a leer un poco de dos tuits; sus identificadores únicos son 499768093921800200 y 499432557906104300.⁹³ Cada uno se identifica de manera única por una cadena de números, que, en este caso, corresponden a dos tuits de agosto de 2014. Muchos materiales digitales, como éstos, pueden parecer frívolos. De hecho, muchos de los ejemplos de cosas como el sitio web de memes en los que me he enfocado en este libro pueden ser generalmente considerados como insignificantes en su superficie. Yo no estoy de acuerdo, pero entiendo de dónde viene ese impulso. Sin embargo, vale la pena enfatizar que hay mucho contenido en plataformas y sistemas que viene con un peso y gravedad visceral mucho más inmediatos. En lugar de elegir algún ejemplo banal de uno de mis propios tuits, he elegido dos ejemplos que se convirtieron en parte de la historia de las protestas de Ferguson. Estos tuits y sus metadatos integrados son documentación de un momento significativo en la historia de Estados Unidos. Más allá de eso, son fuentes primarias (junto con otros en el corpus de tuits compartidos con la etiqueta #Ferguson), que resultan valiosas para comprender tanto el impacto de esta pérdida sin sentido, que no por trágica deja de ser rutinaria, de vidas de personas negras como la aparición del movimiento #BlackLivesMatter.

Comprender los aspectos técnicos de un tuit, como cualquier objeto digital, es esencial para poder decidir cómo trabajar con él en cuanto objeto del patrimonio cultural. Los 115 caracteres que componen el texto del primer tuit son una pequeña porción de lo que en verdad es. Estos tuits ejemplifican muy bien los problemas en torno al esencialismo de pantalla descrito en el segundo capítulo. Para mayor claridad, los tuits son objetos digitales escritos en un conjunto explícito de especificaciones. Se puede acceder a los datos completos de un tuit a través de la “interfaz de programación de aplicaciones” (API) de Twitter desde la cual es posible solicitar un tuit en “notación de objeto de JavaScript” (JSON).

En el caso del tuit 499768093921800200, la respuesta completa en JSON fue de 3 309 caracteres y abarcó 113 líneas. Los caracteres que componen el mensaje del tuit son 3% de todos los datos textuales que vienen en la respuesta. Este tuit, como todos los demás, tiene aproximadamente 97% de metadatos. Tan sólo imagine que el texto de un libro fuera 3% de su contenido

93 Quiero agradecer a Bergis Jules y Ed Summers, quienes, a través de su trabajo con los tuits de Ferguson, me sugirieron éstos entre muchos otros que podría haber incluido aquí; además de este aporte específico, estaré siempre agradecido por todo lo que he aprendido de ambos sobre la preservación digital y las consideraciones interseccionales en torno a ésta.



y el resto estuviera compuesto por metadatos legibles para las máquinas. A partir de esto, se comienzan a vislumbrar algunas de las maneras con profunda diferencias en las que la naturaleza de los objetos digitales debería impactar en nuestro enfoque sobre ellos. Con el fin de demostrar una parte del tipo de información que contiene un tuit, he incluido pedazos seleccionados de estos objetos. Desempacaré paso a paso toda la información contenida en ellos. Me doy cuenta de que los corchetes y otros caracteres son desagradables en términos, pero, por favor, léalo usted mismo; será más fácil que adquiera sentido si lo hace en voz alta.⁹⁴

```

“created_at”: “Thu Aug 14 04:02:53 +0000 2014”,
“id”: 499768093921800200,
“text”: “Don’t Keep much distance from the Police, if you’re close to
        them they can’t tear Gas. To #Ferguson from #Palestine”,
“source”: “<a href =”http://twitter.com” rel=”nofollow”>Twitter Web
        Client</a>”,
“time_zone”: “Jerusalem”,
“ retweet_count “: 534,
“favorite_count”: 268,
“user”: {“id”: 250580585, “name”: “Rajai Abukhalil”, “screen_name”:
        “Rajaiabukhalil”, “location”: “Jerusalem, Falasteen - Berlin”,
        “description”: “رمأل مزل اذا ايسجرن نكو ، تبهذ امنيا كدال ب لمحا”
        Palestinian Doctor & Activist. Co-founder of Physicians for
        Palestine r\nRa...il@gmail.com”, “display_url “:” rajaiabukhalil.
        wordpress.com”, “followers_count”: 3623, “friends_count”: 268,
        “listed_count”: 173, “created_at”: “Fri Feb 11 11:24:43 +0000
        2011”}, }95

```

94 En este caso, he censurado parte de la dirección de correo electrónico de Rajai; aunque este dato es de acceso público, no pensé que fuera aceptable que apareciera en un libro; este punto demuestra aún más el grado en que la información personal puede incrustarse en los objetos digitales publicados abiertamente; además, he incluido este tuit con permiso explícito de Rajai; si bien es un tuit público, él no es una figura de los medios y creo que es importante tener ese respeto.

95 N. del T.: “creado_en”: “Jue Ago14 04:02:53 +0000 2014”, “id”: 499768093921800200, “texto”: “No se alejen demasiado de la policía, si están cerca, no podrán lanzar gas lacrimógeno. Para #Ferguson de #Palestine”, “origen”: “Twitter” “Twitter Web Client”, “zona_horaria”: “Jerusalem”, “cuenta_de_retweets”: 534, “cuenta_de_favoritos”: 268, “usuario”: {“id”: 250580585, “nombre”: “Rajai Abukhalil”, “nombre_de_usuario”: “Rajaiabukhalil”, “ubicación”: “Jerusalem, Falasteen -



Desde el campo de datos creado en la parte superior del tuit, sabemos que se publicó a las 4:02 a. m. del 13 de agosto de 2014. La hora Coordinated Universal Time (UTC) es funcionalmente equivalente a Greenwich Mean Time (GMT). Esto fue cinco días después de que el oficial Darren Wilson disparara y asesinara a Michael Brown en la calle, en Ferguson, Missouri. La zona horaria se establece en Jerusalén, donde eran las 6:02 a. m. El texto del mensaje se encuentra en el campo “texto”: “No se alejen demasiado de la policía, si están cerca, no podrán lanzar gas lacrimógeno para #Ferguson de #Palestina”. Con este tuit, y otros de este y otros usuarios, los activistas en Palestina ofrecieron consejos a los manifestantes y a los activistas en Ferguson. Este tuit y algunos similares se convirtieron en una historia reportada por los principales medios de la prensa. *The Telegraph* señaló, el 15 de agosto, con este tuit incrustado en su artículo: “los palestinos han estado tuiteando consejos sobre cómo lidiar con los gases lacrimógenos a los manifestantes en la ciudad de Ferguson”.⁹⁶ Estos intercambios muestran un replanteamiento de la policía de Ferguson como una fuerza de ocupación a través de su asociación con personas oprimidas del otro lado del planeta.

Los datos adicionales en el tuit proporcionan una variedad de información contextual. Conocemos a su autor, pero no sólo eso, cada tuit viene con datos sustanciales sobre él. En este caso, sabemos que es Rajai Abukhalil. En su biografía, incrustada dentro del objeto que es el tuit, se describe a sí mismo como un médico y activista palestino. También tenemos un enlace a su sitio web. He censurado su dirección de correo electrónico, que en el tuit sí aparece completa. Incluso sabemos el momento en que creó su cuenta (11 de febrero de 2011) y cuántos seguidores tenía el día en que yo solicité el tuit (3 623). Más allá de esta información sobre el autor, también sabemos que el tuit fue enviado desde el cliente web de Twitter, no desde una de las muchas aplicaciones de esta plataforma. El tuit también contiene en su interior una buena cantidad de información que documenta su recepción. Éste fue retuiteado 534 veces y marcado como favorito 268 veces. Es significativo que estos metadatos sobre la recepción y el uso del tuit están vinculados a la fecha en la que lo solicité. Es

Berlin”, “descripción”: “رَمَالا مَزَل اِذَا اُيَسَّجِرَن نَكْفُو ، تَتَبِعْ اَمْنِيَا لِكِدَالِب لِمَح!”, Doctor Palestino & Activista. Co-fundador de Physicians for Palestine \r\nRa...il@gmail.com”, “display_url “:” rajaiabukhalil.wordpress.com”, “cuenta_seguidores”: 3623, “cuenta_amigos”: 268, “cuentas_de_listas”: 173, “creado_en”: “\r\nFeb 11 11:24:43 +0000 2011”, }.

96 Mark Molloy, “Palestinians Tweet Tear Gas Advice to Protesters in Ferguson”, *The Telegraph*, 15 de agosto de 2014, <http://www.telegraph.co.uk/news/worldnews/northamerica/usa/11036190/Palestinians-tweet-tear-gas-advice-to-protesters-in-Ferguson.html>.



completamente posible que otros lo hayan marcado como favorito o retuiteado después, en cuyo caso, este número podría ser diferente si se busca ahora.

En resumen, el tuit contiene en su interior, en forma legible por las máquinas, información sobre su autor, la fecha y el método por el cual fue creado, el texto real del tuit, una breve biografía de su autor e identificadores únicos tanto para el tuit individual como para su autor. Todo esto es sólo la punta del iceberg, pues no he profundizado en los muchos otros campos de datos que hay en cada tuit, pero es más que suficiente para ilustrar hasta que punto los objetos digitales como éstos, en gran medida, se describen a sí mismos. La mayor parte de un tuit está compuesta por metadatos que lo describen.

Al mismo tiempo, para comprender e interpretar este tuit, es necesario saltar a otros puntos de contacto que no se encuentran contenidos en este objeto. Para darle sentido, uno puede visitar el sitio web de Rajai y leer toda la secuencia de tuits que publicó para ver qué otros se compartieron con estos *hashtags* en el mismo momento. Cuando uno ve su tuit incorporado a un artículo o en su línea de tiempo de Twitter, se puede ver su imagen de avatar, pero como objeto textual, el tuit en JSON sólo contiene una URL en la que se puede acceder a esa imagen. Si uno quisiera ver las noticias que se incorporaron o vincularon a este tuit, esa información existe fuera de los límites de este objeto. En lo que respecta a la API de Twitter y Twitter como sistema diseñado a través de su API, un tuit es el conjunto estructurado de datos que se recibe de esta manera.

Otro ejemplo, un tuit del día anterior ofrece más contexto para explorar la información incluida y no incluida en un tuit. Una vez más, es importante que intente leerlo como una cita en bloque, y yo ayudaré a desempacarlo a continuación.

```
“created_at”: “Wed Aug 13 05:49:35 +0000 2014”,
“id”: 499432557906104300,
“text”: “Wow ... A man picks up burning tear gas can and throws it back
        at police. #ferguson pic by @kodacohen @stltoday http://t.co/SASX
        U1yF3E”,
“user_mentions”: [
  {“screen_name”: “kodacohen”, “name”: “Robert Cohen”, “id”: 17073225,
  {“screen_name”: “stltoday”, “name”: “STLtoday”, “id”: 6039302,],
“media_url”: “http://pbs.twimg.com/media/Bu5XQ6KCIAEPh81.jpg”,
“type”: “photo”,
“user”: {“id”: 16661744, “name”: “Lynden Steele”, “screen_name”:
        “manofsteele”, “location”: “St. Louis”, “description”: “Assistant
        Managing Editor for Photography at St. Louis Post-Dispatch, Pulitzer
```



```
Prize winner, insert something funny here”,”created_at”:”Thu Oct 09
03:53:26 +0000 2008”},},
“retweet_count”: 7658,
“favorite_count”: 4697,}”}
```

En este tuit, Lynden Steele, la editora asistente de fotografía del *St. Louis Post-Dispatch*, tuiteó una foto de un hombre que llevaba una camiseta con una bandera estadounidense mientras recogía y arrojaba una lata de gas lacrimógeno. No he mostrado la imagen porque no está en el tuit. Los tuits sólo existen como objetos de texto. En el campo “media_url” se puede ver un enlace a la foto. Sin embargo, si se recopilara este tuit sin tomar por separado la imagen, simplemente se tendría texto sobre la imagen, no la imagen en sí. En el cuerpo del tuit, Lynden citó al fotógrafo; hay que tener en cuenta que Twitter desempaqueta el nombre de pantalla de los usuarios y proporciona su nombre público, en este caso “Robert Cohen”, un fotógrafo para el *St. Louis Dispatch*. El periódico ganaría más adelante un Pulitzer en la categoría Breaking News Photography por este trabajo.⁹⁸

Edward Crawford, el individuo que luego fue identificado como el sujeto de la fotografía, sostiene que, al contrario de lo que decían el mensaje del tuit y el artículo de las noticias, no estaba arrojando la lata de gas a la policía, sino fuera del camino de unos niños. Esa imagen se convirtió en una representación icónica de la lucha en Ferguson. Fue impresa en camisetas y carteles.⁹⁹ Sus cualidades documentales también sirvieron como evidencia

97 N. del T: “creado_en”: “Mie Ago 13 05:49:35 +0000 2014”, “id”: 499432557906104300, “texto”: “Wow... Un hombre recoge una lata de gas lacrimógeno encendida y la arroja de regreso a la policía. #ferguson pic by @kodacohen @stltoday <http://t.co/SASXU1yF3E>”, “menciones_usuarios”: [{“nombre_usuario”: “kodacohen”, “nombre”: “Robert Cohen”, “id”: 17073225, {“nombre_usuario”: “stltoday”, “nombre”: “STLtoday”, “id”: 6039302,}, “media_url”: “<http://pbs.twimg.com/media/Bu5XQ6KCIAEP81.jpg>”, “tipo”: “foto”, “usuario”: {“id”: 16661744, “nombre”: “Lynden Steele”, “nombre_usuario”: “manofsteele”, “ubicación”: “St. Louis”, “descripción”: “Asistente del Editor de Fotografía del St. Louis Post-Dispatch, Ganadora del Premio, inserte algo chistoso aquí”, “creado_en”: “Jue Oct 09 03:53:26 +0000 2008”},}, “cuenta_retweets”: 7658, “cuenta_favoritos”: 4697,}.

98 “The 2015 Pulitzer Prize-Winning Photo graphs from the Post- Dispatch”, *St. Louis Post-Dispatch*, 20 de abril de 2015, <http://www.stltoday.com/news/local/metro/the-pulitzer-prize-winning-photographs-from-the-post-dispatch/collection7a5793c3-9a55-534c-b84a-a27d6d08ef5f.html>.

99 David Hunn, “Subject of Iconic Photo Speaks of Anger, Excitement”, *St. Louis Post-Dispatch*, 24 de agosto de 2014, <http://www.stltoday.com/news/local/crime-and-courts/subject-of-iconic-photo-speaks-of-anger- excitement/article3076e398-2c7b-5706-9856-784c997d0a52.html>.



cuando Crawford fue acusado de asalto e interferencia a un oficial de policía un año después.¹⁰⁰ Este tuit había sido retuiteado 7 658 veces en el momento en que descargué una copia de la API de twitter. Más allá de eso, el tuit en sí estaba insertado en una amplia gama de medios. No sólo documenta este momento en el tiempo, también es el contenido mismo que circuló y sirvió para difundir y construir la narrativa de los eventos a medida que ocurrían en el terreno en Ferguson. El tuit ayudó a hacer la carrera de un hombre blanco y funcionó como evidencia contra un hombre negro. Este último murió de una herida de bala autoinfligida mientras esperaba el juicio. Al respecto, vale la pena mencionar que este ejemplo ofrece una oportunidad para considerar con más profundidad los roles que la documentación y los registros tienen como parte de las estructuras institucionales de opresión y en especial sobre los efectos muy reales que éstos pueden tener en los cuerpos y las vidas de las personas atacadas por esos sistemas opresivos.

Los ejemplos de estos dos tuits presentan un desafío significativo al conocimiento convencional sobre la organización y la descripción de los objetos. Para alguien que quiere documentar y preservar la información sobre un evento, la medida en que estos objetos incluyen cantidades masivas de metadatos es una bendición significativa. Cada tuit contiene en su interior una gran cantidad de metadatos que pueden ayudar a que su contenido sea reconocible. Al mismo tiempo, vemos lo rápido que se rompen los límites entre el contenido de un mensaje y sus descripciones. El objeto, que es el tuit, es más grande por dentro de lo que podríamos haber esperado. Cada tuit está lleno de todo tipo de información sobre sí mismo. De la misma manera, el tuit que aparece en la pantalla como texto acompañado de una imagen en realidad no contiene la imagen. Ningún tuit logra contener gran parte de lo que parece que debería, como las imágenes, audios o videos incrustados.

Podemos pensar en el texto del tuit como sus datos y en todo lo demás como sus metadatos. Sin embargo, el primero también se puede ver como un elemento de una publicación en serie, que es el flujo completo de tuits de un usuario de Twitter. Del mismo modo, cada uno de estos tuits es parte de la secuencia que le apareció a los usuarios de todo el mundo con el *hashtag*

100 Kim Bell, "Ferguson Protester Who Threw Back Tear Gas Canister in Iconic Photo Is Charged", *St. Louis Post-Dispatch*, 26 de agosto de 2015, <http://www.sltoday.com/news/local/crime-and-courts/ferguson-protester-who-threw-back-tear-gas-cannister-in-iconic/article/437076f9-a6a2-5f03-a2fb-d612611a504f.html>.



de Ferguson. Incluso se podría pensar en el texto de estos tuits como anotaciones del momento en que se publicaron. Todos ellos son punteros que señalan a otras cosas en la red, y hacia afuera señalan enfrentamientos muy reales en las calles y a las etéreas redes y flujos del poder cívico, social y económico. Aquí vemos un punto clave sobre los objetos digitales: se describen a sí mismos para ser leídos por una máquina. También se describen entre sí. Más allá de eso, cada bit de metadatos apunta en todas direcciones. Todos los metadatos en un tuit proporcionan contexto al texto del tuit, pero también contextualizan y describen la imagen vinculada e incluso llegan más lejos, pues funcionan como metadatos de incidentes, momentos y flujos de poder.

Esto nos lleva de vuelta al desafío central de los límites de los objetos digitales. En la superficie, parece que un tuit es un objeto útil y significativo para formar parte de una colección. Sin embargo, cuando nos damos cuenta de que éste, tal como es definido por Twitter, no contiene ninguna imagen, parece contrariar nuestra concepción de lo que es un objeto. Por supuesto, a menudo queremos recopilar las imágenes que parecen estar dentro de los tuits, incluso cuando en realidad están solamente vinculadas o metidas a la fuerza en función del momento en que se muestran en el navegador. Del mismo modo, uno podría pensar en cada tuit como una oración de un libro. ¿Tendría sentido recopilar y organizar cada una por separado? En este contexto, como tantos otros, para poder decidir de qué manera recopilar el contenido y, en última instancia, cómo organizarlo y describirlo se requiere un nivel técnico significativo de comprensión sobre la manera en que funciona y ha sido diseñado el sistema subyacente. Aquí nos encontramos de nueva cuenta en un tema del capítulo anterior, que trata sobre el propósito de la preservación. Estos ejemplos ilustran (entre otras cosas) lo fundamental que es comprender cómo se produce el contenido y con qué fin se desea recopilarlo, para que esto pueda realizarse de manera coherente y ética.

A fin de cuentas, la naturaleza de las bases de datos de los nuevos medios, su indexicalidad inherente, implica que este tipo de decisiones sobre la estructura y el orden son mucho más fluidas que para el objeto físico artefactual. En este contexto, tenemos que encontrar la mejor manera de agrupar y procesar los datos para extraer la información significativa que los hará utilizables y descubribles ahora y en el futuro. Al mismo tiempo, necesitamos pensar qué datos deberían o no ser recopilados de acuerdo con la forma en la que podrían terminar siendo utilizados.

La buena noticia es que cosas como los tuits, que están llenos de metadatos extraíbles, se adaptan a la perfección a los principios del enfoque MPLP. Lo



primero que debe hacer es reconocer que los medios digitales vienen con una estructura extensa y a menudo autodescriptiva. Después, sólo déjelos en el orden en el que llegan y explique ese orden para que tanto usted como los demás puedan comprender el alcance y el contexto de lo que tiene. Si desea seguir adelante y recopilar un conjunto de tuits, debe establecer algunos límites para la extensión de una colección determinada y luego debe obtener los datos. Si es hasta cierto punto pequeña, una vez que tiene esto, continuar creando un solo registro para ella. Esto puede funcionar igual que en el caso de la colección 4chan. Es importante explicar un poco de lo que es y de la manera en que está estructurada y dejar que los usuarios la utilicen. Lo anterior deja abierta una amplia gama de modos de acceso potenciales que discutiremos más adelante, en el próximo capítulo. Así, la parte desafiante va más sobre lo que uno debe o no recopilar y cómo proporcionar acceso a ello. Estos puntos se relacionan con las discusiones sobre el desarrollo de colecciones con las que están lidiando proyectos como Documenting the Now, discutido antes.

Ahora volveré a ejemplos aparentemente más banales, como lo que mostré antes de los tuits de Ferguson. Sin embargo, en lo que resta del libro, es vital que considere cómo cualquiera de los ejemplos que analizo se desarrollarían de manera diferente si involucraran a cualquier número de personas cuya información digital pueda ponerlos en riesgo a través de las capas intersectoriales de opresión que estructuran y ordenan a la sociedad. Esto podría incluir publicaciones en Facebook de una persona indocumentada que se encuentra bajo amenaza de deportación, fotos del teléfono de una persona LGBTQ que podrían usarse para exhibirla o acosarla, o cualquier granito de información personal en la red que podría servir como base para cualquier tipo y forma de acoso. Todos los aspectos del trabajo de la preservación, es decir, recopilar, gestionar, organizar, describir y proporcionar acceso a las colecciones, deben abordar los problemas éticos que surgen de este tipo de consideraciones. Retomaré más discusiones sobre estas cuestiones éticas en el próximo capítulo, que trata sobre el acceso.

La multiplicidad de órdenes, variantes y copias

Cuando nos tomamos en serio la idea de que no hay una “primera fila” en las bases de datos como un principio subyacente para comprender los medios digitales, mucho de lo que pensamos sobre la secuencia y el orden de arreglar y organizar las colecciones sale por la ventana. Si bien los sistemas de catalogación siempre han tenido que lidiar con ideas sobre el “orden de estantería” y la disposición y descripción del archivo, se han centrado durante mucho tiempo



en que el “orden original” es el principio. Estas ideas, sin embargo, en un nivel básico, son funcionalmente inapropiadas para trabajar con objetos digitales. Aunque siempre es cierto que en algunos medios subyacentes hay una secuencia lineal de bits que constituye cualquier objeto dado, nuestras interacciones con esos objetos operan a través de una serie de abstracciones.

Los bibliotecarios, los archivistas y los profesionales de los museos aún no se han dado cuenta del potencial que esto les brinda. De manera predeterminada, podemos comenzar asumiendo que hay una multiplicidad de órdenes y posibles órdenes disponibles para nosotros y nuestros usuarios. De la misma manera en que se puede recurrir a todos los archivos y carpetas del escritorio en cualquier momento en función de alguno de los campos de metadatos (alfabético, cronológico, tamaño de archivo, etc.), también podemos trabajar para asegurar que las herramientas y los sistemas que desarrollamos para organizar y proporcionar acceso a las colecciones preserven esa capacidad de manipulación. Esto subraya la resonancia que existe entre las posibilidades de los medios digitales y los objetivos de un enfoque MPLP para la organización y la descripción. Hay que tomar lo dado en el orden en que viene; después se podrá crear cualquier tipo de transformaciones y reordenamientos sin mover los datos subyacentes o el flujo de bits.

La diversidad de formas en las que se puede manipular el contenido digital es liberadora. Hasta cierto punto, podemos dejar que los usuarios finales ordenen y filtren colecciones como tenga más sentido para ellos. Sin embargo, esto significa que las decisiones sobre cómo agrupar y organizar el contenido en conjuntos de información razonables y coherentes son importantes. Volviendo a nuestros ejemplos de Twitter, si usted desarrolla una estrategia para recopilar fragmentos de tuits con base en un criterio temporal, así será la manera sugerida y privilegiada para trabajar con ellos. Si por el contrario usted decide recopilar todos los tuits de un solo usuario a lo largo del tiempo y almacenarlos como una colección, se dará prioridad a esa forma de empacar las cosas, que está centrada en el usuario. Como sea, cualquier persona que esté usando la colección es libre de clasificarla según la cronología, filtrarla por palabra clave o con algún otro número de formas, pero la decisión de lo que está dentro o fuera de una manera determinada de empacar la información tendrá un impacto significativo sobre lo que alguien puede hacer con ella con facilidad. Es decir, se puede requerir que alguien extraiga algún subconjunto de datos de una serie de colecciones diferentes para luego juntarlos y así poder ver o interactuar con ellos de la manera en que lo desea. Esta multiplicidad de órdenes es poderosa, pero también crea nuevos desafíos para dar sentido y relacionar las



variantes de objetos digitales que surgen de la forma en que los usuarios han utilizado las herramientas y los sistemas digitales.

La distinción entre un original y una copia o entre un archivo maestro y sus copias derivadas desempeña un papel central al pensar en la gestión de las colecciones. Es común que las diferencias entre estas relaciones estén atadas a las distintas características que hay entre las concepciones artefactuales e informativas de la identidad de los objetos. A pesar de esto, para muchos objetos analógicos todavía existe, en general, un conjunto más o menos sencillo de relaciones entre originales, variantes, derivados y copias. Éste no es necesariamente el caso de las formas en que manejamos la información digital y tiene algunas implicaciones significativas sobre cómo abordamos la organización y la descripción de los objetos digitales.

A través de una extensa investigación sobre cómo la gente común maneja su información personal digital, la investigadora Catherine Marshall ha demostrado que la facilidad de crear copias y derivados del contenido digital ha provocado una proliferación masiva de variantes de objetos, de los que cada uno tiene su propio sentido distinto pero relacionado.¹⁰¹ De forma ilustrativa, considere las formas en que las personas manejan y trabajan con sus propias colecciones de fotografías digitales.

La vida social de una fotografía digital

Tomo una foto digital con mi teléfono y éste almacena una versión JPEG de la imagen en su panel NAND, también incorpora un conjunto de metadatos técnicos (la fecha y la hora, un código geográfico, datos sobre la marca y el modelo de mi dispositivo, información sobre si se tomó de frente o con la cámara trasera, y nombra el archivo según la secuencia de fotos que he tomado antes). A partir de ahí, podría editar la foto con mucha rapidez (cambiar el contraste, recortarla o rotarla) directo desde mi teléfono y guardarla como una copia adicional o sobrescribir el original.

Podría publicar la foto en Facebook, donde la describiría y podría etiquetar la cara de mi esposa y darle una descripción (usando las funciones de la plataforma). También podría subirla a Instragram, donde la recortaría en un cuadrado y elegiría un filtro que (por cualquier conjunto de razones extrañas)

101 Catherine C. Marshall, “Digital Copies and a Distributed Notion of Reference in Personal Archives”, en *Digital Media: Technological and Social Challenges of the Interactive World*, ed. de Megan Alicia Winget y William Aspray (Lanham: Scarecrow Press, 2011), 89-115.



nos hace creer que hará que nuestras fotos reflejen alguna noción ideal de cómo queremos que nos vean. También tengo mi teléfono configurado para hacer una copia de seguridad automática de todas mis fotos en Flickr: cuando la foto se carga, ese sitio crea de forma automática una serie completa de copias derivadas de la imagen en diferentes tamaños y resoluciones. A medida que las variantes de la imagen se publican con gran velocidad a través de estas diversas plataformas, cada una termina con diferentes títulos y subtítulos, fechas y horas de publicación, etiquetas y metadatos adicionales, y cada una acumula sus propios metadatos sociales (me gusta, favoritos, comentarios). Muy rápidamente, a través de las prácticas cotidianas de trabajar con fotos en mi teléfono, ahora hay docenas de copias de esta imagen. ¿Cómo se podrían relacionar y administrar estas versiones de imágenes?

El resultado neto de esta situación es que nuestro mundo de objetos digitales implica una gran cantidad de variantes y derivados principales y leves. En la investigación de Marshall, se ilustraron situaciones en las que podrían existir variantes de un video musical animado en una docena de sitios diferentes con metadatos distintos y una gran cantidad de otras transformaciones (por ejemplo, diferentes tipos de compresión, transcodificación, introducciones, etc.). Ampliando mi ejemplo de las fotos personales, a medida que las personas manejan esta información y al final la exportan, la respaldan y la transfieren a otras organizaciones, estas variantes existirán en una multiplicidad de formas diferentes: un directorio con todas mis fotos que respaldo cada año, una copia de mi archivo de Facebook, una copia de mi archivo de Flickr. Si bien es trivial identificar copias exactas mediante la comparación de la información sobre la integridad de los archivos, también es probable que la mayoría de estas variantes no se registren como idénticas, de acuerdo con un control de integridad, ya que son al menos un poco diferentes. Con ese fin, hay una gama de técnicas desarrolladas en otros campos para identificar computacionalmente copias cercanas y variantes (analizando de manera visual fotos para detectar similitud con otras fotos, por poner un ejemplo).

Si bien es posible gastar mucho tiempo y energía seleccionando y clasificando estas relaciones y tomando decisiones sobre qué copias deben conservarse, de acuerdo con los principios de MPLP, es mejor conservar muchas de esas variantes y permanecer abierto a explorar métodos computacionales para relacionarlas y conectarlas mejor en el futuro. Si alguien, por alguna razón, adquiriera todas mis fotos digitales, podría ser razonable tomar una decisión en el más alto nivel de la organización para mantener un conjunto y no otro, por ejemplo, mantener fotos del archivo de Facebook y descartar la copia de



seguridad de Flickr. La sugerencia de Marshall es que hagamos las paces con el hecho de que vamos a lidiar con un rango de copias idénticas o cercanas con una variación significativa y que probablemente lo que debemos hacer es comenzar a desarrollar enfoques para armonizarlas. Desde esta perspectiva, pasamos a una noción distribuida de estos objetos, en la que nos centramos menos en qué versión es la mejor o la “maestra” y más en desarrollar enfoques para automatizar las relaciones entre copias y copias cercanas y en crear métodos para describir estos agregados y grupos de objetos vinculados. Hay una gama de técnicas emergentes para este tipo de coincidencias difusas, basadas en el análisis computacional de las imágenes representadas, que son bastante prometedoras. Si bien la tecnología no está aquí para hacer eso todavía, es suficiente para mantener la adición de copias de forma que sean utilizables ahora y permitan esperar el desarrollo de estos métodos y técnicas futuras.

Con la proliferación de contenido interrelacionado y variado también han surgido nuevos métodos para retroceder un poco y permitir que diferentes comunidades identifiquen e interpreten sus propios materiales. Explicaré de manera breve algunos ejemplos de cómo esto puede funcionar y del impacto que puede tener cuando se ordena y describe el contenido.

Relegar la descripción a los colaboradores o a las comunidades de usuarios

Empoderar a los colaboradores y a las comunidades de usuarios para que ellos sean quienes describan y cataloguen los objetos se está volviendo cada vez más popular. Esto tiene la ventaja de representar el contenido en los mismos términos en que lo utilizaría un usuario y también descarga el trabajo y la mano de obra necesarios para describir las cantidades cada vez mayores de contenido que las organizaciones desean recopilar. Compartir tres ejemplos muy diferentes de cómo se desarrolla esto en sistemas de colección relacionados pero muy distintos (identificación y descripción de los GIF animados para una exhibición de museo, un depósito institucional y una aplicación móvil para recopilar historia oral) ayudará a demostrar algunas de las ventajas de esta técnica. A medida que los bibliotecarios, archivistas y curadores se trasladan a este tipo de trabajo, necesitan convertirse en profesionales reflexivos del diseño de la experiencia del usuario. En este proceso, nos convertimos menos en productores de descripción y más en facilitadores de su producción.



Los GIF de reacción como una forma gestual

El formato de intercambio de gráficos, conocido como GIF, fue creado en 1987 por CompuServe. Es uno entre muchos formatos de imagen,¹⁰² pero también es de los pocos que ha llegado a tener un significado cultural tan específico. Los GIF permiten reproducir animaciones en bucle, lo que extrañamente los convierte en una especie de imagen en movimiento. Como consecuencia, el nombre de este formato se ha vuelto un término cultural aplicado a los bucles animados, y los llamados “GIF de reacción” se han convertido en una forma cultural propia. En general, estos últimos son animados y muestran un cuerpo en movimiento, a menudo de una película o de un programa de televisión, que las personas usan en el discurso en línea para expresar una reacción. Por lo tanto, podría tratarse de un bucle corto de la actriz Jennifer Lawrence asintiendo y diciendo “OK” sarcásticamente mientras muestra su pulgar hacia arriba o una animación de Michael Jackson en el video musical de *Thriller* comiendo palomitas de maíz.

En 2014, Jason Eppink, curador del Museum of the Moving Image, realizó un proyecto con el objetivo de crear una colección de estos GIF para documentar su función cultural y su uso.¹⁰³ Su trabajo en la instalación y exposición en línea, *The Reaction GIF: Moving Image as Gesture*, ofrece la oportunidad de pensar en un enfoque para organizar y describir el contenido digital que sea útil para desempacar el potencial del medio.

Mientras observaba la forma en que las personas usaban los GIF de reacción en el sitio de redes sociales Reddit, Eppink se dio cuenta de que había mucha comprensión compartida de lo que éstos significaban en las comunidades en línea. Al verlos como un medio a través del cual la gente insertaba una forma gestual en su discurso, el curador decidió desarrollar una colección de GIF y exhibirlos en el museo. Para decidir cuáles utilizar, recurrió a la comunidad de “redditors”, con el propósito de que le ayudaran a identificar lo que era, al menos en ese momento, el canon de los GIF y luego para que describieran sus significados. El resultado fue una colección de imágenes y explicaciones de lo que esos GIF representaban para los usuarios de dicha comunidad. Al acudir

102 Para conocer más sobre la historia de los GIF animados, ver Jason Eppink, “A Brief History of the GIF (So Far)”, *Journal of Visual Culture* 13, núm. 3 (2014): 298-306, DOI: 10.1177/1470412914553365.

103 Para más información sobre esta colección, ver Owens, “Exhibiting .Gifs: An Interview with Curator Jason Eppink”, *The Signal* (blog), 2 de junio de 2014, <http://blogs.loc.gov/digitalpreservation/2014/06/exhibiting-gifs-an-interview-with-curator-jason-eppink/>.



a los redditors para identificar los GIF más populares y luego pedirles que los interpretaran, Eppink logró seleccionar tanto elementos relevantes como mostrar el significado de este lenguaje digital vernáculo desde el punto de vista de los participantes.

Esto en sí mismo sirve como documentación de un tipo de lenguaje vernáculo digital, la amplia gama de variaciones de estos diferentes GIF, pero el tipo de documentación derivada de la discusión entre los redditors también ofreció un medio para seleccionar estos objetos y describirlos. Los redditors identificaron los GIF de reacciones que consideraban los más utilizados y conocidos y se involucraron en la discusión para definir lo que querían decir.

Eppink partió de un enfoque vernáculo/folclórico para estudiar estos GIF. En este contexto, no le interesaba saber quién era el autor, sino lo que el GIF significaba para las diversas personas comunes que lo usaban en línea. Del mismo modo, sería un poco difícil tratar de identificar al autor del signo “OK” que se hace con la mano, pero no sería particularmente extraño estudiar el significado contextual que existe en torno al uso de ese gesto. Entonces, si bien es posible rastrear a través del tiempo e identificar cuándo y dónde surgió el primer GIF animado de Michael Jackson comiendo palomitas de maíz, para el propósito de esta colección es más importante el hecho de que ese GIF particular es icónico. Como ejemplo están los “GIFS de comer palomitas” que a menudo se publican en los hilos de discusión en los que se quiere expresar diversión y entretenimiento de acuerdo con el debate o drama que se desarrolla dentro de una discusión.¹⁰⁴ El recurrir a una comunidad para participar en la descripción y la selección de una colección es en sí mismo una herramienta poderosa para resolver algunos desafíos. También vale la pena subrayar que, si el enfoque es la documentación cultural, muchos de los tipos de metadatos predeterminados (autor, datos creados, editor, etc.) pueden no ser las formas más útiles de abordar la creación u organización de una colección. Además, usar este marco de significado cultural para comprender las colecciones se ajusta de forma natural a la variedad de maneras en que los sociólogos, antropólogos, folcloristas y académicos de la comunicación han abordado el estudio de las formas culturales. Así, el adoptar este enfoque, cuando uno puede anticipar que ese tipo de académicos serán los usuarios, le permitirá generar una mejor alineación entre su enfoque de organización y descripción y las formas en que éstos piensan sobre ese contenido.

104 “Popcorn GIFs”, Know Your Meme, acceso el 28 de abril de 2021, <http://knowyourmeme.com/memes/popcorn-gifs>.



Un repositorio institucional

Cada vez más instituciones de educación superior tienen repositorios. Si bien esto puede sonar como una función genérica de una organización, en este caso, ha llegado a significar un conjunto muy específico de características para gestionar, preservar y proporcionar el acceso a sus resultados académicos. Muchos de estos sistemas son administrados por la biblioteca de una institución. Por ejemplo, cuando terminé mi tesis, la deposité en el Mason Archival Repository System (MARS), que es el repositorio institucional de la George Mason University (GMU).

Depositar mi tesis aquí no fue diferente a subir una foto a Flickr. Me dieron una cuenta de usuario, inicié sesión, cargué el PDF, completé una serie de formularios web donde puse el título, elegí el departamento del que me estaba graduando dentro de un menú desplegable y metí mi resumen. El sistema grabó la fecha en que lo subí. Marqué una casilla para que se pudiera acceder a mi tesis de inmediato, y así estuvo disponible en línea. Usted la puede encontrar en la sección College of Education and Human Development de la colección electrónica de tesis y disertaciones de la GMU. Me imagino que en el pasado un catalogador habría necesitado crear un registro, pero ahora el sistema está configurado para que ese trabajo sea transferido al autor. Los repositorios institucionales se utilizan para mucho más que tesis y disertaciones: contienen cosas como actas de reuniones, pódcast, presentaciones de PowerPoint, conjuntos de datos de investigación, grabaciones de video, etc.

Las fortalezas de este tipo de sistemas (como es, en gran medida, que el volante haya sido entregado a los usuarios) son a la vez su mayor debilidad. Es fácil que dichos sistemas se conviertan en un “hotel de segunda clase”, donde las personas suben de manera inconsistente y describen inadecuadamente los objetos digitales.¹⁰⁵ Asimismo, tienden a funcionar bien para algo como tesis y disertaciones, donde es posible insistir en que alguien no pueda graduarse sin depositar su trabajo. Sin embargo, más allá de esto, también tienden a representar de manera desigual los registros de una institución. Es decir, la extensión de este tipo de colecciones vivirá o morirá en función de lo bien que se incentive la participación de sus usuarios. Cuando una biblioteca contrata a alguien para describir y catalogar de forma sistemática alguna colección, se obtienen metadatos y niveles de descripción con un grado alto de consistencia. Cuando

105 Uno de los mejores textos sobre la extensión de los límites y retos de establecer un repositorio digital funcional sigue siendo el de Dorothea Salo, “Innkeeper at the Roach Motel”, *Library Trends* 57, núm. 2 (2008): 98-123.



se entrega esa función para que cada usuario final de un sistema lo haga, puede terminar fácilmente con datos irregulares y escasos. El mismo conjunto de problemas surge en contextos muy diferentes.

StoryCorp.Me: colaboración abierta o crowdsourcing de la historia oral

Es probable que muchos lectores estén familiarizados con StoryCorps, los fragmentos de entrevistas de historia oral que se comparten cada semana en la National Public Radio. Desde 2003, StoryCorp ha recopilado más de 50 mil de estos testimonios, los cuales han sido recopilados en cabinas de audio que viajan por todo el país. Las entrevistas son descritas y almacenadas por el personal y los archivistas de StoryCorps, y al final se archivan para ser preservadas en la Library of Congress. En 2015, StoryCorps diseñó y lanzó StoryCorp.me, una aplicación móvil que permite a personas de todo el mundo realizar entrevistas de StoryCorp usando su teléfono, así como ingresar metadatos para ellas y subirlas para incluirlas como parte del archivo. Como usted se puede imaginar, esto expandió rápida y dramáticamente el tamaño de la colección. En poco tiempo StoryCorp obtuvo un número mayor de entrevistas provenientes de la aplicación que todos los testimonios recopilados desde las cabinas de audio durante más de una década. Debo enfatizar que este caso no es único. Muchas organizaciones utilizan cosas como la plataforma para la publicación de colecciones de código abierto llamada Omeka para crear estas colecciones a partir del *crowdsourcing*. Las aplicaciones que recopilaban entrevistas y otros esfuerzos de colaboración colectiva vienen con el mismo conjunto de problemas potenciales que los materiales de un repositorio institucional. Es decir, como los metadatos de esta colección son generados por el usuario, parece inevitable que resulten más inconsistentes e incompletos que los que han sido creados en el pasado por profesionales capacitados.

Entonces, el diseño de la experiencia del usuario, a través de pruebas iterativas del desarrollo de interfaces y flujos de trabajo, se vuelve crítico para las capacidades de este tipo de sistemas. Esto termina siendo una especie de proyecto de investigación en ciencias sociales. ¿Cómo se motiva o incentiva a las personas a participar en estos sistemas? ¿Qué tipos de campos de metadatos se deben usar? ¿Se deberían utilizar vocabularios controlados y menús despleables? ¿Cuándo es necesario tener un campo de texto libre? ¿Cuántos campos de metadatos se le puede pedir a alguien que complete? ¿Se deben establecer restricciones sobre los tipos de archivos permitidos para subir al sistema? ¿Se validan los archivos que se suben y, si hace falta información, se obliga a los usuarios a volverlos a enviar? ¿Cuántos campos es probable que alguien



complete? ¿Qué campos deberían ser obligatorios y cuáles opcionales? Éstas son preguntas que le conciernen al campo de investigación de las ciencias sociales y del desarrollo de la interacción humano-computadora, y no tenemos todavía una respuesta para ellas. Debido al desarrollo de este tipo de sistemas para construir colecciones digitales, ahora estas interrogantes se han convertido en puntos centrales de la biblioteconomía y del trabajo de las instituciones del patrimonio cultural. Estas últimas deben invertir tiempo, recursos y personal para tener éxito en la producción de interfaces de recopilación comprensivas, consistentes y utilizables. Las decisiones de diseño en cuanto a cómo estructurar las aportaciones e interacciones de los usuarios dejarán una marca duradera en el contenido y en la estructura de esas colecciones.

Las descripciones vinculadas y en la red

Muchos objetos digitales se indexan, describen y anotan entre sí. Considere, por ejemplo, todos los enlaces que aparecen en los artículos publicados en *Drudge Report*. El hecho de que el diario digital se vincule a los sitios de esos enlaces nos dice algo sobre ellos. Volviendo a los ejemplos de los metadatos de un tuit individual, si usted tomara todos los tuits que un político importante ha hecho a lo largo del tiempo, terminaría con un conjunto coherente de tuits, pero también con una serie de anotaciones de las URL mencionadas en ellos. Es decir, si usted tomara cada uno de los tuits de esa colección que incluyen alguna URL en específico y todos los tuits de otras colecciones que mencionan esa misma URL, obtendría mucha información que describiría, contextualizaría y comentaría el recurso que sea que resida en la URL. Este conjunto de conexiones vinculadas se convierte en una poderosa forma de contexto.

Los alcances de esto se relacionan con temas del capítulo anterior, que habla del desarrollo de colecciones. Por ejemplo, varios proyectos web que se dedican a la recopilación han utilizado un conjunto de información como medio para identificar el contenido que se debe seleccionar. La British National Library desarrolló TwitterVane, una herramienta de código abierto para recopilar las URL mencionadas en los tuits que aluden a un término determinado.¹⁰⁶ Del mismo modo, Internet Archive creó una colección de archivos web utilizando las URL extraídas de los 13 millones de tuits que mencionaron el *hashtag*

106 Para una discusión más detallada sobre el proyecto TwitterVane y el trabajo relacionado con él, ver Ian Milligan, Nick Ruest y Jimmy Lin, “Content Selection and Curation for Web Archiving: The Gatekeepers vs. the Masses”, *Digital Libraries (JCDL), 2016 IEEE/ACM Conference on Digital Libraries* (2016), 107-110, <http://ieeexplore.ieee.org/abstract/document/7559571/>.



de Ferguson en las protestas después del asesinato de Michael Brown. En estos casos, una colección digital se convierte en la base para determinar el alcance de otra, y las dos, funcionalmente, se glosan y sirven de contexto entre sí.

Las técnicas de procesamiento del lenguaje natural, es decir, los métodos computacionales para identificar patrones en textos se utilizan cada vez más para mostrar las relaciones entre los objetos y los metadatos descriptivos. Existen potentes cajas de herramientas de código abierto para la extracción de entidades por nombre (un enfoque destinado a la identificación de nombres de personas y de lugares, fechas y horas expresados en una variedad de formas textuales), y los investigadores están trabajando en la aplicación de estas técnicas para ayudar a la descripción de los objetos. A medida que se perfeccionen, será posible que un bibliotecario o un archivista las lleve a cabo sobre un corpus de texto o que aplique las técnicas visuales o de texto hablado sobre varias colecciones de imágenes o de audio, para así vincular, etiquetar y clasificar el contenido según las sugerencias proporcionadas por estas tecnologías computacionales.

Es apropiado clasificar casi todos los objetos digitales como datos en sí mismos y también como metadatos que describen otros objetos. Para este fin, debemos pensar en “la descripción” y “lo descrito” como un límite difuso. El futuro implicará descubrir cómo aprovechar el hecho de que los medios digitales cada vez establecen más referencias entre sí. Encontrar la manera de mostrar y documentar estos enlaces de forma automatizada proporcionará una gran cantidad de metadatos descriptivos potencialmente valiosos. Es probable que también sea la única forma en que podamos trabajar en la escala del contenido digital.

De catalogador a *data wrangler*¹⁰⁷

Las posibilidades de acción de los objetos digitales presentan oportunidades para repensar cómo los organizamos y los describimos. La naturaleza de las bases de datos de la información digital hace que algunos aspectos de la organización sean mucho más directos. Podemos permitir a los usuarios finales asumir mucha más injerencia en el filtrado y la clasificación del contenido de una manera que les sea útil en un momento dado. Al mismo tiempo, las formas en que recopilamos y fragmentamos el contenido pueden privilegiar unas maneras de abordarlo sobre otras. Sigue siendo crítico considerar cómo

107 N. del T: el término *data wrangler* es relativamente nuevo y, por el momento, no se encontró una traducción al español, por lo que en este libro se emplea la palabra en inglés.



se organiza la información digital, pero el énfasis está menos sobre el objeto individual y más sobre la creación de adiciones útiles.

Por fortuna, la lógica de la informática insiste en que existan diferentes niveles de metadatos incrustados en los objetos digitales. En cierto nivel, todos estos objetos se autodescriben, ya que vienen con nombres y extensiones de archivo, marcas de tiempo, etc. Todos esos metadatos son lo que permite a los profesionales del patrimonio cultural y a sus usuarios manipularlos y trabajarlos rápidamente. El desorden de los objetos digitales, es decir, sus copias variantes, el rizoma de enlaces e interrelaciones, proporciona una gran cantidad de elementos para resolver las relaciones de red que existen entre ellos. En este contexto, establecer los límites de los objetos digitales en relación con el propósito de la preservación se vuelve primordial.

Estamos dejando atrás un mundo en el que un archivero o un catalogador establecen un orden y redactan una descripción, para entrar a otro donde estos profesionales influyen, reúnen y dan sentido a los flujos de información. Se trata menos de imponer arreglos o de aplicar descripciones y más sobre traerlas a la superficie y de aclarar y decidir qué orden dentro del contenido digital será privilegiado. En muchos casos, los roles también están cambiando para permitir que varios tipos de usuarios describan y organicen el contenido. En este espacio, se vuelve cada vez más importante tomar las lecciones del enfoque MPLP de la teoría archivística y aplicarlo a las prácticas que nos permiten trabajar en niveles más altos de organización para dejar que los aspectos de nivel más bajo de ordenamiento y descripción sean cubiertos por los metadatos incrustados y las formas de orden y estructura que vienen por principio con todo tipo de objetos digitales.

Capítulo 8

La habilitación de los usos y los accesos multimodales

GeoCities, el servicio gratuito de alojamiento web y una de las primeras comunidades virtuales, cerró abruptamente en el 2009. Por fortuna, el Internet Archive y una banda de archivistas rebeldes llamados Archive Team hicieron un esfuerzo para archivar el sitio. El Archive Team permitió el acceso a un *terabyte* de datos del sitio archivado y el Internet Archive dispuso la entrada al sitio a través de su WayBack Machine. De esa copia masiva de datos del Archive Team surgieron una serie de interfaces importantes y significativas. Además, de esa información en crudo, algunos artistas también crearon múltiples proyectos, como el *Deleted City*, que visualiza el contenido del sitio a través de diagramas de árbol, y *One Terabyte of the Kilobyte Age*, que representaba e interpretaba páginas del sitio en Tumblr. Esta historia, en la que profundizaré más adelante, ilustra una serie de temas clave para el uso de las colecciones digitales. Su acceso rápido y en masa puede facultar a los usuarios para crear otras formas de acceder a ellas. El futuro del acceso digital comienza con el acceso total y masivo que otorga poder a los usuarios y avanza hacia una gama de proyectos *boutique* justificados por las necesidades particulares de cada comunidad.

A fin de cuentas, la preservación se trata de proveer acceso en el futuro, pero hacerlo en la actualidad también es importante. Es difícil justificar la colección y preservación de cosas si no son de valor para las personas involucradas. Al trabajar en procurar el acceso al contenido digital, es fundamental que usted se asegure de que las maneras en las que éste se adquiere, organiza y proporciona sean útiles e inteligibles para sus usuarios. Mientras más pronto proporcione el acceso, más rápido sabrá si el trabajo funciona y tiene sentido



para ellos. Es decir, al subir el contenido y ponerlo a disposición de los usuarios, usted puede comenzar a incorporar las contribuciones que ellos hagan para refinar su enfoque y seguir adelante.

Este capítulo examina algunas de las oportunidades y los desafíos encontrados al proporcionar el acceso a las colecciones digitales. Existen cosas simples que cualquiera puede hacer para proveer el acceso inmediato al contenido hasta cierto nivel. Es fundamental pensar muy bien en las cuestiones legales y éticas involucradas en procurar el acceso a las colecciones. Estas consideraciones dan entrada a una serie de cuestiones clave sobre qué tipo de restricciones de acceso tienen sentido en diferentes contextos. Con esto en mente, trabajaré con una variedad de ejemplos de algunos métodos y enfoques para proporcionar el acceso a colecciones digitales específicas, con las que mostraré la importancia de tener un enfoque multimodal para el acceso. Esto permite a sus usuarios ofrecer retroalimentación sobre qué haría que el contenido fuera más útil para ellos. También abre la oportunidad para que los usuarios tomen la batuta y diseñen interfaces para su contenido de manera que otras personas también lo puedan utilizar.

En conjunto, los ejemplos de este capítulo demostrarán que el acceso es y continuará siendo cada vez más multimodal. Debido a las posibilidades inherentes de la información digital, es lógico que veamos una variedad de ensamblajes, adiciones, interfaces y métodos para el acceso y uso de las colecciones digitales. En este contexto, tiene sentido pensar en “las maneras”, en lugar de “una sola manera”, en las que se puede proporcionar el acceso al contenido. Esto involucra que pensemos cada vez más de manera transversal en un espectro de modos de acceso. Uno de sus extremos es al por mayor: métodos para proporcionar el acceso masivo a las colecciones como conjuntos de datos y de manera uniforme a todo el contenido. En el otro extremo está la *boutique*: conjuntos de materiales con curadurías especializadas, interfaces especiales personalizables o modos de remediar o destacar el contenido para casos particulares de uso.

Dado que existen múltiples formas en las que uno se puede enfocar para dar acceso a un conjunto determinado de objetos, por lo general, es buena idea empezar por lo que es más fácil o sencillo y explorar otros enfoques conforme el tiempo y los recursos lo permitan. Como ha sido el caso a lo largo del libro, la consideración clave para la toma de decisiones sobre qué métodos de acceso tienen sentido en un contexto dado implica entender el costo-beneficio de cada enfoque para respaldar las necesidades de los usuarios o los participantes.



Comenzar por lo simple y dar prioridad al acceso

Cuando en una biblioteca de investigación un usuario pide ver un libro en un idioma desconocido, por lo general, el bibliotecario lo traerá y dejará que lo vea. Quizá dicho bibliotecario no tenga idea de lo que dice el texto, lo que no le impide saber cómo dar acceso a él. En esta situación, se asume que un investigador que solicita un libro como ese necesita tener las habilidades necesarias para darle sentido. En el nivel más básico, podemos proporcionar esta clase de acceso para cualquier objeto digital que estemos preservando. Aquí está: tómalo. Volviendo al principio del enfoque MPLP descrito en el capítulo anterior, tendríamos que estar trabajando en proveer este nivel de acceso a cada cosa que hayamos adquirido tan rápido como sea posible. Debería tomar muy poco tiempo convertir las cosas recopiladas por nosotros en cosas a las que proporcionamos el acceso, al menos en una forma mínima. El nivel de acceso básico tendría que tomar días, semanas o meses, no años.

Una vez que usted haya adquirido un objeto o colección, debería ofrecer una mínima descripción acerca de ellos y proporcionar tanto acceso abierto como le esté permitido y sea ético. El ejemplo del archivo 4chan en Stanford en el capítulo pasado ejemplifica este mínimo requerido. Los bibliotecarios de Stanford reciben los documentos; crean una colección a nivel de registro; evalúan si pueden compartirlos abiertamente en la red; publican la colección o el elemento en la web, donde cualquiera puede descargar una copia o verificar que su copia sea fiel al registro proporcionado por Stanford. La colección fue evaluada, seleccionada, descrita y se hizo accesible a nivel de colección en cuestión de horas.

Las posibilidades de los medios digitales ofrecen un potencial significativo para el acceso y el uso del contenido digital. Nuestra experiencia con el consumidor de *software* puede interponerse en comenzar a proporcionar el acceso al contenido digital. Muchos bibliotecarios y archivistas no proveen el acceso a un conjunto de contenidos hasta que tienen en marcha un sistema sofisticado que lo reproduce en una interfaz fácil de usar. Vale la pena notar que “amigable” es en sí mismo un marco ideológico para el cómputo que a menudo entorpece una apreciación crítica de los medios computacionales.¹⁰⁸ Tenemos que superar el deseo de tener esta clase de interfaces para todo, en las que sólo hacen falta unos cuantos clics sobre un objeto digital para que “funcione”. Recordemos, esto es parte de la problemática suposición del

108 Al respecto, ver Emerson, *Reading Writing Interfaces: From the Digital to the Bookbound* (Mineápolis: University of Minnesota Press, 2014), capítulo 2.



esencialismo de pantalla. No existe una interfaz ni un medio de acceso inherente para cualquier objeto digital. En su lugar, existen varios modos de reproducir dicho contenido, que pueden o no importarle a un usuario particular en un caso de uso específico.

Esto significa que los profesionales de la preservación digital, para empezar, tienen que estar de acuerdo con decir “Aquí está: tómallo”. Incluso si no saben cómo abrir determinado archivo, los usuarios pueden explorarlo en un editor hexadecimal y leer los metadatos embebidos en él. En el nivel más básico, la norma tendría que ser proporcionar el acceso a los materiales digitales en el sitio. Con esa consideración, se requiere pensar un poco más en las posibilidades de acceso que proporciona la red abierta.

La tiranía del sistema de descubrimiento

Antes de abordar las cuestiones que giran en torno a las restricciones del acceso, primero quiero abordar el tema de las interfaces y los sistemas de descubrimiento. Para quienes trabajan en instituciones del patrimonio cultural, la solución al problema del acceso es lo que sea que su sistema ofrezca. De modo que si usted realiza un proyecto de formación de una colección usando Omeka, éste le permitirá proporcionar el acceso a cada elemento mediante su interfaz. De forma similar, si usted usa DSpace como sistema de repositorio, entonces incluirá una interfaz para el usuario final. En cualquier caso, use lo que esté a su alcance. Sin embargo, no permita que estos sistemas determinen cuál podría ser el mejor modo de proporcionar el acceso a su contenido.

No deje que la funcionalidad de un sistema sea el principal motor de las decisiones que tome sobre la descripción y la organización de su contenido. Volviendo al ejemplo del archivo 4chan, la página de dicha colección en el repositorio de Stanford se parece mucho a la de un objeto, pero en realidad es una descripción de la colección entera. Hoy en día, cualquier sistema de descubrimiento en uso es temporal. El novedoso sistema de hoy es la plataforma vieja y frustrante de mañana. Y usted necesitará migrar su contenido de ahí hacia otro lugar. En tres años o en una década habrá un nuevo sistema, y usted no debe tomar decisiones significativas respecto a lo que hará a largo plazo con base en el conjunto de funciones que tiene una aplicación en este momento. Entonces, parece más simple publicar en línea el contenido que se tiene, ya sea en directorios o en páginas HTML muy sencillas, y después integrar eso con cualquiera de los sistemas que de cualquier manera se usarían.

Para seguir con la idea del acceso multimodal, si a usted se le proporciona un sistema de descubrimiento con el cual trabajar, éste debe servir como el



punto de referencia y el modo de acceso inicial. De cualquier forma, no piense que ésta es la única manera en que se usará el contenido. Se trata simplemente de un medio para un fin en el presente. Dicho esto, regresaré a la reflexión sobre las oportunidades de acceso fuera de cualquier sistema o infraestructura particulares.

Las posibilidades y los problemas de la red abierta

La red es una plataforma de acceso maravillosa. El contenido publicado en Internet se vuelve accesible a nivel global casi de inmediato. Parece difícil medir qué tan masivo puede ser el impacto que esto ha tenido y sigue teniendo en cuanto al acceso al patrimonio cultural material.

Un archivista puede tener una colección de diez mil documentos de procesadores de texto, convertirlos en lote a HTML y subirlos a la red con una estructura de vínculos simple para que un motor de búsqueda los indexe. De esta manera, se obtiene el acceso y el descubrimiento global del contenido de una forma veloz. En esta etapa, si alguien hace una búsqueda en Google con el nombre de una persona desconocida, puede llegar de forma directa al documento número 3 587 en el que se le menciona.

La cualidad de ser descubierto que tiene el contenido es transformadora. Este tema habría sido abordado por la ciencia ficción, si no fuera porque la red lo ha vuelto parte de nuestra vida cotidiana. De cualquier modo, esta capacidad de descubrimiento requiere que seamos mucho más reflexivos de lo que hemos sido en el pasado respecto a los efectos que puede tener el acceso inmediato a la información. Antes, el nombre de una persona presente en una página de un documento en un archivo habría sido accesible para cualquiera que, primero, de alguna forma supiera buscar en dicha colección, aunque el nombre ni siquiera estuviera descrito en los metadatos y, segundo, que después fuese de forma física al archivo y repasara todo el material para poder encontrarlo. El esfuerzo que se requería para obtener acceso a algo era una forma de privacidad sustentada en el oscurantismo. Como resultado, es imperativo que quienes administran las colecciones digitales sean más conscientes acerca de las problemáticas éticas y legales que emergen de este tipo de forma de acceso y descubrimiento transformacional.

La ética, la privacidad, los derechos y las restricciones

Consignas como “la información quiere ser libre” parecen tener una verdad aparente. Sin embargo, hay tres razones principales por las cuales quienes



trabajan para proporcionar el acceso a las colecciones digitales no pueden simplemente tomar este tipo de frases como su mantra: el derecho de autor, la privacidad y el respeto a las normas culturales. Describiré de manera breve cada uno de estos puntos y después, a lo largo del capítulo, demostraré con ejemplos de qué manera funcionan los diferentes acercamientos al acceso para abordar estas cuestiones legales y éticas.

Casi todo el trabajo que se produce está protegido por los derechos de autor. Yo no soy abogado. Esto no es un libro sobre derechos de autor, así que no tome estos comentarios breves como una asesoría jurídica. En lugar de eso, puede usar mis palabras como una incitación a la reflexión sobre estas cuestiones y a la profundización para hacer una investigación o una consulta legal propias. Los derechos de autor conceden permisos exclusivos de distribución al creador de una obra. En los Estados Unidos, las bibliotecas tienen un conjunto de excepciones establecidas en la sección 108 de la ley sobre derechos de autor que trata de la preservación y el acceso a las colecciones. Estas excepciones son por desgracia anticuadas para tratar con la información, pero siguen siendo importantes.¹⁰⁹ Además, hay estipulaciones de uso justo que las bibliotecas están utilizando cada vez más para apoyar sus enfoques respecto a procurar el acceso a sus materiales.¹¹⁰ Además de estos derechos, las instituciones culturales a menudo hacen acuerdos con los donadores del contenido o con aquellos de quienes lo adquieren para establecer qué permisos otorga el titular de los derechos a la entidad para distribuir su contenido. Asimismo, el creciente movimiento a favor del uso y reuso abierto del trabajo creativo a través de iniciativas como las licencias Creative Commons ha creado una variedad de posibilidades para coleccionar, preservar y proporcionar el acceso al contenido.

Más allá de los derechos de autor, hay cuestiones éticas sobre la privacidad que se ponen en juego al hacer que un contenido esté ampliamente disponible. Por ejemplo, los registros administrativos de una organización pueden incluir información sobre los salarios de cada uno de sus empleados, sus números de seguridad social e incluso datos de sus cuentas bancarias. Toda esa informa-

109 Para una revisión de las cuestiones relativas a la sección 108 y los objetos digitales, ver Section 108 Study Group, *The Section 108 Study Group Report* (Washington DC: The Library of Congress, 2008), <http://www.section108.gov/docs/Sec108StudyGroupReport.pdf>. El informe tiene casi una década, pero todavía no ha dado lugar a una revisión sustancial de la ley.

110 Para una introducción práctica y útil respecto a cómo las bibliotecas utilizan el uso justo, ver Prudence S. Adler *et al.*, *Code of Best Practices in Fair Use for Academic and Research Libraries* ([s. l.]: Association of Research Libraries 2012), <http://www.arl.org/storage/documents/publications/code-of-best-practices-fair-use.pdf>.



ción privada podría usarse para causar daño a estos individuos. Muchos tipos de datos para la investigación plantean cuestiones éticas similares. Otro caso son los investigadores de medicina y ciencia social que trabajan en instituciones de educación superior y que a menudo recopilan datos confidenciales acerca de la orientación sexual de la gente o su consumo de drogas. Si esta información llegara a hacerse pública, podría ser peligroso para los participantes de la investigación. De la misma forma, un académico que estudia los movimientos sociales y el uso de las redes sociales de los activistas podría terminar recopilando información de interés para los individuos que trabajan para suprimir dichos movimientos o incluso para oprimir todavía más a esas personas. Más allá del terreno de los datos personales, incluso existen cuestiones éticas respecto a la información confidencial. Es posible que un zoólogo que estudia una especie en peligro de extinción recopile información sobre ubicaciones de animales que los cazadores furtivos podrían utilizar para realizar su cometido, con lo que pondría a la especie en riesgo de un peligro mayor. Éstos son sólo algunos ejemplos, pero todos subrayan que el ignorar los asuntos de la privacidad de la información puede provocar un daño real.

Las instituciones del patrimonio cultural tienen la responsabilidad ética de trabajar para asegurar que sus compromisos con la preservación y el acceso también estén alineados con una responsabilidad social compartida. Esto requiere analizar con detenimiento los enfoques de acceso restringido al contenido que toman en consideración este tipo de problemáticas. En consonancia con la perspectiva interactiva de estos últimos capítulos del libro, las cuestiones éticas que se relacionan con el acceso muchas veces nos llevan a regresar a las primeras decisiones sobre qué coleccionar, cómo hacerlo y cómo procesar dichas colecciones. Muchas veces, la mejor manera de lidiar con este tema es censurar, editar, o bien, garantizar que la información que pueda ser peligrosa para los individuos simplemente no se recopile ni se guarde. Si usted, de hecho, cuenta con información que pueda ocasionar un daño verdadero, debe saber que destruirla es el modo más seguro para proteger a aquellos que estarían en peligro. Así que cualquier decisión para conservar la información sensible debe llegar como resultado de una valoración premeditada respecto al beneficio y el daño potencial que podrían derivar de su conservación.

Junto con los problemas de la privacidad, existe otra categoría de preocupaciones éticas relacionadas con esto. Los diferentes grupos culturales y étnicos alrededor del mundo tienen distintas normas y reglas respecto a lo que se puede y se debe compartir con alguien en determinada situación o escenario. Ésta es una cuestión particularmente significativa relacionada con las colecciones



de los materiales culturales pertenecientes a comunidades nativas. En términos históricos, los poderes coloniales han ignorado las cuestiones culturales. Por ejemplo, podría haber objetos digitalizados a los cuales sólo deberían tener acceso los integrantes de una familia, o que sólo deberían usarse o verse en una temporada o época del año en particular, o que sean considerados sagrados y sólo deban emplearse de un modo específico. Kimberly Christen, de la Washington State University, dirige un equipo que desarrolla y da respaldo a Mukurtu, un sistema de administración de colecciones de código abierto que se enfoca en devolver el control de las decisiones sobre el acceso a las comunidades nativas.¹¹¹ De modo similar, iniciativas como Traditional Knowledge Labels están intentando desarrollar formas de hacer comprensibles esta clase de restricciones de acceso.¹¹² Estos ejemplos nos recuerdan que el objetivo de las instituciones del patrimonio cultural no debe ser el acceso sólo por el simple acceso. Éstas deben funcionar como un recurso para las comunidades, no como sus cosechadores ni saqueadores. La historia nos enseña que muchas veces no ha sido así. Las bibliotecas, los archivos y los museos han servido y, por defecto, siguen sirviendo, en muchos casos, como infraestructuras del colonialismo y la opresión.¹¹³ Tenemos que hacer un mejor trabajo al respecto.

Vale la pena subrayar que, aunque mucho del trabajo que se enfoca en estas normas culturales emerge del contacto con comunidades nativas, sus implicaciones son de gran alcance. Al tiempo que las instituciones del patrimonio cultural trabajan para volverse más socialmente responsables, es necesario que reconozcan que los flujos de poder muchas veces son los responsables de su propia existencia. Como consecuencia de estas estructuras de poder, una porción significativa del trabajo que las bibliotecas, los archivos y los museos han realizado a lo largo del tiempo en el desarrollo de las colecciones ha funcionado como parte de sistemas de control y opresión. Muchas instituciones públicas y no lucrativas surgieron de las colecciones personales de los ricos y continúan construyendo colecciones en continuidad con la dirección establecida por sus progenitores. Así que es preciso pensar tanto en los asuntos relativos

111 Kimberly Christen, “Working Together: Archival Challenges and Digital Solutions in Aboriginal Australia”, *SAA Archaeological Recorder* 8, núm. 2 (2008): 21-24.

112 Jane Anderson y Kimberly Christen, ““Chuck a Copyright on It”: Dilemmas of Digital Return and the Possibilities for Traditional Knowledge Licenses and Labels”, *Museum Anthropology Review* 7, núms. 1-2 (2013): 105-126.

113 Jarrett M. Drake, “Expanding #ArchivesForBlackLives to Traditional Archival Repositories”, *On Archivy*, 27 de junio de 2016, <https://medium.com/on-archivy/expanding-archivesfor-blacklives-to-traditional-archival-repositories-b88641e2daf6#.6w6jkmgul>.



a las normas culturales como en las cuestiones de la privacidad individual en una amplia gama de situaciones vinculadas a las colecciones. En particular, es importante considerar estos temas cuando se trabaja con iniciativas enfocadas en documentar o preservar los registros de las comunidades que han sido o están siendo oprimidas. Una forma de ayudar es hacer un esfuerzo para incorporar sistemáticamente las voces y las perspectivas de los oprimidos en cada aspecto de la recopilación, la preservación y la tarea de proporcionar el acceso a la cultura.

Estas tres áreas, derechos de autor, privacidad y normas culturales, son consideraciones generales que se deben tomar en cuenta mientras se desarrollan los planes para hacer accesible cualquier colección. De acuerdo con el enfoque interactivo de la parte práctica de este libro, vale la pena subrayar que el acceso no es algo que se deba dejar para más tarde. Usted debe pensar en los modos de acceso desde el principio. Para cada una de estas cuestiones legales y éticas, el modo más directo de confrontarlas es restringir el acceso al contenido. Eso podría implicar la administración de un archivo oscuro, al que no se permite el acceso, o el dar acceso restringido a una colección con permisos exclusivos para las personas que visiten una sala de lectura o sólo para quienes crean una cuenta y firman algún acuerdo sobre lo que harán y no harán con el contenido. Dicho esto, hay formas más sofisticadas de abordar el tema del acceso, tanto en los casos en que existen restricciones significativas como donde no las hay. Explorar varios ejemplos ayudará a demostrar estos puntos.

Los niveles multimodales de acceso y restricción

Existen formas cada vez más sofisticadas de proporcionar el acceso de manera útil a los aspectos de una colección y respetar al mismo tiempo las restricciones de acceso que surgen de los derechos de autor, la privacidad y las normas culturales. El mejor modo de ilustrar esto es mediante algunos ejemplos concretos. En esta sección, presento tres colecciones diferentes de material digital: la National Software Reference Library, los corpus de Google Books y HathiTrust, y los CD-ROM de Theresa Duncan de Rhizome. Las restricciones al acceso en cada uno de estos casos tienen más que ver con los derechos de autor que con la privacidad o con las normas culturales, pero sus enfoques son ilustrativos de los métodos que podrían usarse para estas dos áreas. Varios de ellos ilustran el valor de crear y compartir datos derivados de colecciones. Éste es un punto que exploraré más a fondo en la próxima sección, que se enfoca en el potencial de la remediación, los derivados y las mejoras para permitir el acceso y el uso de las colecciones.



El acceso a un archivo oscuro: la National Software Reference Library

Creada en el año 2000, la National Software Reference Library (NSRL), una división del National Institute of Standards and Technology (NIST) del Departamento de Comercio de Estados Unidos, contiene una colección masiva de archivos de paquetes comerciales de *software* (desde QuickBooks, Microsoft Word y AutoCAD hasta *World of Warcraft*).¹¹⁴ Hasta el momento en que escribo este libro, esta lista incluía alrededor de 140 millones de archivos digitales. La mayor parte del *software* de esta colección se compró, pero otro tanto se adquirió por donación, razón por la cual la NSRL no tiene el derecho de distribución de los archivos. Se trata casi en su totalidad de *software* comercial que se vende y se utiliza en la actualidad.

Esta peculiar colección se creó con el objetivo de estandarizar las prácticas de la identificación de *software* como parte de las investigaciones forenses de cómputo. Es decir, uno quiere ser capaz de reconocer de forma rápida qué parte del contenido de una computadora es *software* estándar y qué parte es material potencialmente único y relevante para convertirse en el foco de una investigación. Como se discutió antes en relación a la integridad, una de las características clave de los objetos digitales es que es posible generar valores criptográficos *hash* (MD5, SHA-1, etc.) para identificar un archivo de forma única. Así, al centralizar una biblioteca masiva de *software* comercial, la NSRL puede proporcionar valores *hash* para todos los archivos que contiene, los cuales pueden ser utilizados por un número indefinido de organizaciones para identificar concluyentemente qué *software* está en cualquier dispositivo de cómputo.

Si bien todo el contenido de la colección está protegido por derechos de autor, la NSRL sí puede distribuir sus metadatos. Éstos son útiles porque contienen información descriptiva que es clave (título, editorial, etc.), además de los valores *hash* para cada uno de los archivos asociados con cada programa de cómputo individual. Aunada a la información derivada, la NSRL mantiene una copia de la colección para fines de investigación. Aunque no se permite a los investigadores el acceso a la colección *in situ*, la NSRL acepta hacer consultas y búsquedas en el corpus masivo de archivos únicos. En ese sentido, este corpus de archivos únicos ha jugado un papel clave como conjunto de datos de investigación a partir del cual se han desarrollado varios métodos y enfoques de las ciencias computacionales.

114 Para más detalles sobre cómo trabaja la NSRL, ver Owens, “Life-Saving: The National Software Reference Library”, *The Signal* (blog), 4 de mayo de 2012, <http://blogs.loc.gov/thesignal/2012/05/life-saving-the-national-software-reference-library/>.



La NSRL funciona como un archivo oscuro. La colección de *software* se mantiene en una *air-gapped network*, es decir, en una red separada que no está conectada a las computadoras fuera de las salas del NIST, donde se encuentra la biblioteca. Por esta razón, no sólo es imposible acceder a dichos archivos desde el exterior, sino que incluso una persona que se encuentra dentro del NIST tampoco puede hacerlo sin entrar de forma física a ese espacio específico. No es muy distinto al Svalbard Global Seed Vault, el repositorio global de semillas de último recurso. Gracias a la derivación, los metadatos de la colección y el publicar y compartir esa información, el contenido de la NSRL es muy utilizado. Mientras que los archivos reales se preservan y permanecen inaccesibles, la información derivada de ellos tiene un uso inmediato para una variedad de comunidades interesadas en ellos. De forma similar, al permitir el análisis computacional de los contenidos de sus colecciones, el NIST proporciona una manera valiosa para que otros los aprovechen. Mientras que la mayoría de los lectores de este libro no contarán con la sofisticada infraestructura que tiene el NSRL, existen algunas lecciones que se pueden aprender de este caso. La publicación de los metadatos y de la información descriptiva derivada computacionalmente sobre cualquier colección dada puede ser muy útil. En particular, la unicidad de los valores *hash* ofrece un modo potencial para identificar copias exactas en varias colecciones.

El uso no consuntivo: HathiTrust y Google Ngram

HathiTrust es una asociación de más de 120 instituciones de investigación y bibliotecas que proporcionan acceso en línea a millones de obras digitalizadas. En particular, contiene más de siete millones de libros digitales. La sociedad se originó a través del proyecto Google Books, en el que Google se asoció con algunas bibliotecas para digitalizar los libros de sus colecciones. Muchas de las obras digitalizadas en este proceso siguen bajo la protección de los derechos de autor, por lo que no pueden compartirse de manera amplia y libre. En este sentido, cabe aclarar que dentro de la red de socios de HathiTrust es posible hacer uso de las obras de la colección dentro de lo estipulado en la ley de derechos de autor, a lo que se ha llamado comúnmente “uso no consuntivo”. Éste es un término un poco confuso que, en general, se refiere al análisis computacional de textos como un conjunto de datos en oposición a una lectura directa del material. De forma similar a la colección del NSRL, HathiTrust dirige un centro de investigación que permite a los académicos realizar análisis computacionales y minería de datos al corpus completo de libros para fines de este tipo de estudios. También dirige iniciativas para identificar las obras huérfanas que hay



dentro de la colección y trabaja para clarificar el estatus de sus derechos para hacerlas más accesibles. Como tal, este esfuerzo proporciona valor añadido y significativo al trabajo de proveer una mayor apertura y acceso a obras cuyos derechos de autor resultan difíciles de determinar.

Un ejemplo de los tipos de información derivada que ha sido muy útil, en particular en una variedad de contextos de investigación, es el corpus de n-gramas de Google y su visualizador. El Google Ngram tiene como base un subconjunto remediado y procesado de un corpus de libros digitalizados dentro del proyecto Google Books para que cualquier persona interesada en ello estudie las tendencias históricas del uso de las palabras a lo largo del tiempo. A través de este tipo de interfaz se puede estudiar cualquier cantidad de cambios sociales y culturales, tal como se manifiestan en los textos de los libros. También es posible descargar los datos subyacentes. El conjunto de datos que alimenta al Ngram Viewer contiene todas las palabras de las obras, pero lo que muestra es la frecuencia con la que aparecen y no las narrativas en sí. Por esta razón, los datos digitales compartidos ahí representan la información de las obras y no propiamente su contenido, pues está protegido por los derechos de autor.

Las máquinas virtuales remotas: los CD-ROM de Theresa Duncan

La diseñadora Theresa Duncan lanzó, a mediados y finales de la década los 90, una trilogía de juegos de narración interactiva para el sistema operativo de Mac (*Chop Suey* en 1995, *Smarty* en 1996 y *Zero Zero* en 1997). Estos juegos son interesantes en cuanto ejemplos de la historia de la narración interactiva y como documentación de algunos de los tipos de trabajos dirigidos por mujeres. Sacar a la luz el recorrido de la historia del liderazgo femenino en los videojuegos es importante para contrarrestar la tradición en este ámbito que a menudo ha borrado, ignorado o rechazado las contribuciones de las mujeres a los juegos y la cultura del juego. Con el paso del tiempo, incluso si se tiene una copia del trabajo de Duncan, se ha vuelto cada vez más difícil poderlo jugar debido a que representa un desafío conseguir una máquina vieja o poner en marcha un sistema de emulación o virtualización para usarlos. De esta manera, se ha vuelto más complicado tener acceso a ellos.

Rhizome asumió la tarea de preservar y proporcionar acceso a estas obras.¹¹⁵ Con el uso del bwFLA Emulation como la plataforma de servicio de

¹¹⁵ Para más detalles sobre cómo funcionó este proyecto, ver Dragan Espenschied *et al.*, “(Re-) Publication of Preserved, Interactive Content—Theresa Duncan CD-ROMs: Visionary Video-



código abierto, se puso en marcha un sistema donde cualquiera puede visitar su sitio y conectarse a una máquina virtual que corre el sistema operativo requerido e incluye una copia de cada uno de los juegos. En realidad, se trata de una experiencia bastante notable: hay que dar clic en el vínculo y, tras un momento, se carga una copia del sistema Mac OS antiguo en el navegador. Éste es el mismo tipo de acceso que se proporciona para las *laptops* de Salman Rushdie archivadas en el Emory que se describieron antes.

Es importante mencionar que no se trata de una distribución de copias del juego. Más bien, un usuario accede de forma remota a una copia del juego, la cual corre en un servidor alojado por Amazon. A los jugadores se les proporciona acceso al entorno del juego, pero los archivos reales están corriendo en la computadora remota. Mientras que un usuario podría crear capturas de pantalla, no podría realmente copiar los archivos subyacentes del juego para acceder a ellos después. Los usuarios pueden interactuar con el juego, pero sólo en la medida en que la computadora remota les proporciona acceso. Aunque no es un requerimiento fundamental para estas obras en particular, este tipo de enfoque de acceso remoto a entornos computacionales virtualizados o emulados ofrece un potencial significativo para dar accesibilidad al contenido sin distribuirlo abiertamente y que se pueda seguir difundiendo.

La remediación, los derivados y las mejoras

Cuando hablo de “acceso” en este libro, por lo general me refiero a proporcionar información digital a los usuarios potenciales. Con esto en mente, cada vez más instituciones del patrimonio cultural están estableciendo relaciones mucho más ricas y complejas con sus usuarios, lo que deriva en diversas formas de uso, reuso y remediación del contenido. El resto de este capítulo explora una serie de ejemplos que demuestran algunos de estos ecosistemas emergentes de acceso multimodal.

En algunos casos, este tipo de interacciones permite a los usuarios desarrollar nuevos métodos para proporcionar el acceso a los materiales de una colección. Muchas de las nuevas formas de investigación digital se han convertido en interfaces de colecciones digitales. Esto complica el modelo general de la relación que hay entre los usuarios y los proveedores de las colecciones, ya que,

games for Girls”, en *Proceedings of the 12th International Conference on Digital Preservation* (Chapel Hill: School of Information and Library Science, University of North Carolina at Chapel Hill, 2015), 233-234, <http://phaidra.univie.ac.at/o:429524>.



de manera creciente, los académicos están produciendo formas derivadas de las colecciones que más tarde tienen sus propias audiencias y usuarios, quienes las utilizan para fines diversos. En muchos de estos casos, la institución dedicada al patrimonio cultural ofrece el acceso masivo a la información “al por mayor” y después los investigadores hacen la mayor parte del trabajo para que la información le sea útil a las audiencias.

Por el contrario, las instituciones dedicadas al patrimonio cultural invierten una cantidad considerable de tiempo y de recursos en mejorar el uso de estas colecciones. Las discusiones anteriores sobre *crowdsourcing* ofrecen puntos relacionados con esto e invitan a considerar la coproducción de la adquisición, la organización y la descripción de las colecciones, pero muchos de estos casos demuestran cómo ese trabajo está vinculado a la evolución emergente del acceso. A la par de proveer formas más directas de acceso a los elementos de las colecciones, las instituciones dedicadas al patrimonio cultural también han facilitado la producción de interpretaciones y de recursos de tipo derivativo. Esto incluye desde bibliografías críticas, exposiciones y catálogos hasta ediciones documentales y críticas. Estas mejoras de los objetos y las fuentes permiten modos más profundos de comprometerse a proporcionar el acceso a los materiales. En oposición al modelo de “al por mayor” antes mencionado, el enfoque “*boutique*” requiere una justificación sustancial de las razones por las que vale la pena invertir tiempo y recursos en crear este tipo de formas derivadas o mejoradas del contenido.

Estos puntos se ilustran mejor a través de la discusión de cinco ejemplos: las interfaces de GeoCities, el acceso masivo a los periódicos digitalizados, el uso del reconocimiento facial para desarrollar interfaces para registros de archivos, la remediación de los libros electrónicos del dominio público y la creación de recursos que permitan que las colecciones sean útiles para los profesores de educación básica. Juntos, estos casos demuestran diversos niveles de compromiso y diferentes tipos de ecosistemas que han surgido alrededor del uso y reuso de las colecciones digitales. Al mirarlos todos, se pueden trazar los cada vez más diversos métodos y estrategias en las que las instituciones deben pensar en cuanto al acceso al contenido digital.

Los múltiples puntos de entrada: los archivos de GeoCities

Como mencioné antes, GeoCities fue un servicio de alojamiento web gratuito y una comunidad virtual temprana. Con base en una metáfora, cada usuario seleccionaba en qué ciudad estaría localizado su sitio web, lo que serviría luego como un punto clave en la navegación del sitio. Algunos elementos eran



geográficos y otros tópicos. GeoCities se lanzó en 1994, y Yahoo! lo compró en 1999. Diez años después, en 2009, se cerró la sección de Estados Unidos, la cual tenía más de 38 millones de páginas de contenido.

Dado el lugar prominente que tuvo GeoCities en la historia de la red, su cierre provocó una serie de esfuerzos para preservarlo. Buena parte del sitio había sido archivado a lo largo del tiempo por el Internet Archive a través de sus actividades archivísticas regulares. Además, el Internet Archive se involucró en el almacenamiento focalizado del sitio poco antes de su cierre. Al mismo tiempo, un movimiento comunitario liderado por el archivista Jason Scott y el Archive Team, entre otros actores, también se dedicó a copiar partes del sitio. Un año después de la clausura, el Archive Team publicó un archivo *torrent* (un medio para compartir archivos entre pares) que incluía cerca de un *terabyte* de información del sitio archivado.¹¹⁶

En esta ocasión, el Archive Team hizo un solo vertido masivo de datos, el cual dio origen a una variedad de proyectos de creación de interfaces significativas e importantes para dicha información. Dos obras derivadas, creadas a partir de la publicación de este volcado de datos, demuestran cómo los usuarios de las colecciones digitales cada vez desempeñan roles que corresponden más una forma híbrida de desarrollo de interfaces y de creación de interpretaciones.

The Deleted City es una pieza creada por el artista Richard Vijgen, la cual interpreta y presenta una interfaz para mostrar información.¹¹⁷ Se trata de una interfaz de pantalla táctil que permite explorar todo el contenido de GeoCities. También utiliza una interfaz de diagramas de árbol en la que el tamaño de las diferentes áreas del sitio está representado de forma visual según el número de archivos que contiene. Pensada como una obra de arte, ha adquirido su propio valor y se ha expuesto como instalación en galerías, y ahora también está disponible en la red. Dicho esto, *The Deleted City* proporciona acceso a los contenidos del sitio a nivel del archivo de una forma fácilmente navegable. Esta obra de arte es un modo muy útil para explorar y entender los contenidos de esta colección en particular.

Los folcloristas digitales Olia Lialina y Dragan Espenschied crearon *One Terabyte of the Kilobyte Age*, otra obra derivada de este conjunto de datos, pero

116 Jason Scott, "Archiveteam! The Geocities Torrent", *ascii* by Jason Scott (blog), 27 de octubre de 2010, <http://ascii.textfiles.com/archives/2720>.

117 Richard Vijgen, *Deleted Cities* [2011], Rhizome, acceso el 28 de abril de 2021, <http://rhizome.org/art/artbase/artwork/deleted-cities/>.



anclada en un enfoque muy distinto. Como una “recreación” a GeoCities, esta pieza funciona tanto para interpretar como para proporcionar acceso a los contenidos del archivo.

One Terabyte of the Kilobyte Age toma las páginas individuales de GeoCities y genera una forma para representarlas con fidelidad en navegadores y sistemas operativos viejos, incluso con la resolución de pantalla que habría sido común para ese entonces. Tras conseguir este nivel de autenticidad en la reproducción, se toman capturas de las páginas y luego se publican en el Tumblr del proyecto. Al convertir las páginas en imágenes y ponerlas en Tumblr, se difunden en los muros de una amplia gama de usuarios de esta red social. Muchos de ellos luego vuelven a compartir las imágenes en sus muros. Unas cuantas imágenes de los sitios de GeoCities se publican en el *feed* de Tumblr cada hora. Hasta ahora, se han publicado más de 90 mil imágenes a través de este proyecto, el cual tiene su propio valor como obra creativa, pero también es una forma de publicar el contenido del archivo para compartir la colección de manera más amplia y para hacerla accesible.

El Internet Archive ofrece otras maneras de interactuar con las copias archivadas del sitio mediante GifCities. Cualquiera puede usar la Wayback Machine, el método general del Internet Archive para obtener acceso a cada uno de los sitios archivados a partir de sus URL. Además, el Internet Archive creó una interfaz de búsqueda personalizada para la colección GifCities. Ésta permite a los usuarios buscar por palabra clave a través de todos los GIF animados archivados en él. Parte de la vernácula digital de GeoCities fue su abundancia en cuanto a los primeros GIF animados de la red. El equipo de Internet Archive se dio cuenta de que al proporcionar una forma especializada de buscar y explorar esos GIF, probablemente, generaría un interés nuevo y adicional en la colección. El lanzamiento de la interfaz de búsqueda de GifCities desvió una cantidad considerable de atención hacia la colección en varios blogs de tecnología y en la prensa. Aunque esos GIF siempre estuvieron allí para que cualquiera pudiera encontrarlos, la creación de la interfaz de búsqueda personalizada generó un nuevo interés en los usuarios.

A través de estos tres ejemplos podemos ver cómo surge una serie de patrones. Al proporcionar acceso masivo a una copia archivada de los datos, Archive Team abrió la oportunidad para que diferentes artistas usaran los contenidos de la colección y crearan obras que se sostienen por sí mismas como proyectos creativos, pero que también existen en un continuo funcionamiento como interfaces de la información archivada y como formas de exposición y publicación de interpretaciones de esos datos. Al mismo tiempo, la estrategia del Inter-



net Archive de desarrollar una interfaz adicional, GifCities, para una parte de la información en su archivo, estimuló su uso y fomento una interacción adicional. Esta multiplicidad de modos de acceso, que abarca desde los datos masivos hasta las interfaces a nivel de elemento, nace de la cadena de estructuras integradas y de los metadatos que son inherentes a los objetos digitales nativos.

Los datos masivos que los usuarios quieren: la información del National Digital Newspaper Program

El National Digital Newspaper Program (NDNP) de la Library of Congress de Estados Unidos publica una colección masiva de páginas digitalizadas de periódicos estadounidenses a través del sitio *Chronicling America*. Hasta este momento, incluye más de 11.5 millones de imágenes y es el resultado de un enfoque bastante particular y distribuido en cuanto al desarrollo de colecciones. El National Endowment for the Humanities financia la digitalización del contenido de los diarios a través de subsidios para las instituciones dedicadas al patrimonio cultural en los Estados Unidos que tienen copias en microfilm de series importantes de periódicos de sus áreas. Los fondos para las subvenciones de estos proyectos apoyan tanto la digitalización como el proceso mediante el cual expertos en la materia ayudan a seleccionar los periódicos que son más significativos en cada región para priorizar su reproducción digital. La Library of Congress archiva las copias en microfilm, digitalizadas de acuerdo con sus estándares, y luego las publica en la red desde *Chronicling America*.

Este sitio web es un recurso maravilloso para los investigadores. Además de que pueden navegar dentro de la colección, gracias a que las páginas pasaron a través del “reconocimiento óptico de caracteres” (OCR), también pueden hacer búsquedas de palabras clave por todo el corpus. La búsqueda de texto completo sigue siendo un medio increíblemente transformativo para proporcionar el acceso a los materiales que forman parte de una colección. A partir de ello, los usuarios pueden leer las páginas digitalizadas individualmente con sus términos de búsqueda resaltados. *Chronicling America* también proporciona otros modos de acceso adicionales: una API, que permite a los programadores interactuar con la información de la colección; métodos para obtener acceso masivo a las páginas digitalizadas, y descargas masivas de los datos OCR. A través de la API, los desarrolladores e investigadores pueden realizar consultas personalizadas dentro del corpus de datos.

En la práctica, la API ha sido menos utilizada que otros formatos. En contraste, los conjuntos de datos masivos del OCR se han usado ampliamente. David Brunton, uno de los principales desarrolladores del sitio, brinda algunas



explicaciones sobre este fenómeno que son muy relevantes para pensar en el acceso a colecciones como conjuntos de datos.¹¹⁸ Las lecciones aprendidas a través de este proyecto ofrecen conocimiento respecto a qué métodos de acceso multimodal tendría más sentido priorizar en casos particulares. El uso de la API requiere un entendimiento sofisticado de programación web. La mayoría de los usuarios potenciales de los datos no han sido desarrolladores web y tampoco tienen acceso a recursos en esa materia. En cambio, la mayor parte de los académicos interesados en los datos masivos a menudo son investigadores de humanidades con poca experiencia en programación. Entonces, la descarga de la totalidad de las imágenes de la colección, o incluso sólo una selección, proporciona con rapidez una cantidad masiva de datos a dichos usuarios, pero en realidad no es posible utilizarlas de forma significativa. También vale la pena señalar que, en este momento, no hay muchos enfoques o métodos para trabajar a nivel computacional con las imágenes como objetos de investigación, por lo que este tipo de acceso no se conecta con facilidad con los métodos que los usuarios podrían usar.

Por el contrario, los datos masivos del OCR son relativamente pequeños. En vez de descargar todos los archivos de imágenes grandes, el OCR sólo contiene el texto de cada página en un tamaño mucho más compacto. Así, el texto del OCR es algo con lo que un usuario puede trabajar en una *laptop* de calidad alta pero modesta. Aunado a ello, existe una variedad de herramientas gratuitas y de código abierto para manejar, en términos computacionales, información textual. De esta manera, los datos del OCR, derivados de las imágenes digitalizadas y creados con fines de investigación, lo que los hace propensos al error y a contener “ruido”, son los más utilizados de entre estas formas masivas de proporcionar acceso.

Como ha sido el caso en varios de los otros ejemplos sobre el acceso a las colecciones discutidos hasta ahora, muchos de los académicos que trabajan con estos datos han producido formas propias de estudio digital que ofrecen diferentes interfaces para interactuar y explorar partes de las colecciones que manejan. Por ejemplo, el proyecto Viral Texts de la Northeastern University ha investigado y trazado las redes de reimpresión de los periódicos y las revistas en el siglo XIX.¹¹⁹ A través de este trabajo, se han producido publicaciones

118 David Brunton, “Using Data from Historic Newspapers”, *The Signal* (blog), 5 de septiembre de 2017, <https://blogs.loc.gov/thesignal/2017/09/using-data-from-historic-newspapers/>.

119 Para más información acerca del trabajo del equipo de Viral Texts, ver Ryan Cordell y Abby Mullen, “‘Fugitive Verses’: The Circulation of Poems in Nineteenth-Century American Newspa-



académicas tradicionales, como artículos de revistas, así como interfaces en línea vinculadas a textos que aparecen repetidos en diferentes periódicos con enlaces a sus fuentes. De forma similar, los historiadores de Virginia Tech utilizan los datos del NDNP para rastrear y analizar la cobertura mediática del brote de influenza de 1918. Una vez más, su trabajo produjo publicaciones académicas, conjuntos de datos derivados e interfaces en línea para dar acceso a esa información.¹²⁰

Aquí emerge un patrón: proporcionar vías para el acceso masivo a las colecciones permite la creación de formas derivadas adicionales de información. Es importante mencionar que en este caso el OCR era en sí mismo una forma derivada de información de la colección. Debido a que ésta encajaba mejor con los recursos y herramientas que los usuarios potenciales ya tenían a la mano, fue fundamental que la Library of Congress cumpliera con sus necesidades a tal punto que ellos y sus herramientas pudieran trabajar mejor y ofrecer ese tipo de información.

Nos dirigimos al establecimiento de un diálogo cada vez más complejo con los usuarios, que estará basado en el desarrollo continuo de métodos y estrategias novedosas para trabajar computacionalmente con los objetos digitales. Por ejemplo, tomando el caso de las páginas de periódicos, hace poco un equipo de académicos en humanidades, dirigido por la bibliotecaria Elizabeth Lorang en la University of Nebraska, ha desarrollado estrategias para identificar el contenido poético de un corpus.¹²¹ Esto ha requerido del uso de técnicas computacionales de visión para analizar la estructura visual de las imágenes. El acceso masivo a los archivos de imagen se convierte en una parte crucial de este tipo de trabajos. Por supuesto que éstos sólo son posibles gracias a, primero, el desarrollo de las herramientas de cómputo de visión que realizan esta clase de análisis y, segundo, el creciente desarrollo de las herramientas que vuelven, hasta cierto punto, más fáciles estas investigaciones, más intensivas en su aspecto computacional. En ambos puntos se encuentran rasgos que con toda seguridad seguirán impulsando este diálogo continuo entre los tipos de acceso que las instituciones pueden ofrecer y la habilidad de las distintas comunidades de usuarios para utilizar dichos métodos. Conforme

pers”, *American Periodicals* 27, núm. 1 (2017), <http://viraltxts.org/2016/04/08/fugitive-verses/>.

120 Neil Fraistat, *Shared Horizons: Data, BioMedicine, and the Digital Humanities* (College Park: Digital Repository, University of Maryland, 2013), <http://drum.lib.umd.edu/handle/1903/14721>.

121 Elizabeth M. Lorang *et al.*, “Developing an Image- Based Classifier for Detecting Poetic Content in Historic Newspaper Collections”, *D-Lib Magazine* 21, núms. 7-8 (2015), <http://digitcommons.unl.edu/libraryscience/340/>.



las tecnologías sigan avanzando, los usuarios serán capaces de desarrollar formas adicionales de trabajar y hacer uso de nuestras colecciones.

Algo que no debemos perder de vista es que proporcionar el acceso al contenido es una labor que jamás estará terminada. Las nuevas técnicas y métodos para trabajar con objetos digitales requerirán siempre estar reconsiderando qué tipo de presentaciones y conjuntos de datos derivados podrían ser útiles, lo que probablemente generará nuevos y adicionales modos e interfaces para la manipulación y el acceso al material de una colección.

La inversión de las relaciones de control: la verdadera cara de White Australia

A comienzos del siglo xx, Australia estableció una serie de leyes llamadas White Australia Policy, que se enfocaban en limitar la migración de los habitantes no europeos al país. Estas políticas racistas continuaron vigentes durante las décadas de 1950 a 1970. En los National Archives of Australia existe documentación extensa sobre cómo y cuándo estos grupos de migrantes fueron controlados y oprimidos. A través del proyecto Invisible Australians, los historiadores Kate Bagnall y Tim Sherratt se han abocado a convertir estos documentos en un medio para sacar a la luz las historias y las narrativas de las personas a las que, en principio, buscaban controlar esas leyes.¹²² Los resultados de este trabajo ofrecen bastantes modelos con potencial para que las instituciones dedicadas al patrimonio cultural puedan desarrollar y cultivar las relaciones que tienen con sus usuarios, con el objetivo de crear modos de acceso y uso para sus colecciones.

El producto principal de este trabajo es una interfaz que presenta un muro de caras bajo el título “El verdadero rostro de Australia blanca”. Los usuarios pueden hacer clic en cualquiera de las imágenes para ver el documento del que se extrajo cada rostro. La mayor parte de los documentos provienen de las colecciones digitalizadas de los National Archives of Australia que están en línea. Sherratt y Bagnall seleccionaron una serie de archivos, descargaron las imágenes de cada uno y, luego, utilizaron un *software* de reconocimiento facial para identificar las caras incrustadas en las imágenes adjuntas a los documentos. Desde esta página, un usuario puede desplazarse hacia abajo por un muro aparentemente interminable de retratos de hombres, mujeres y niños que fueron registrados como control del gobierno. Aunque mirar los documentos refuerza la narrativa de dominio y control, extraer los rostros de los que sufrieron esa opresión desplaza el énfasis hacia sus historias de vida.

122 Kate Bagnall, “Invisible Australians”, *Asian Currents* (abril/mayo de 2012): 14-15.



El proyecto de Sherratt y Bagnall ha sido objeto del reconocimiento internacional y ha dado lugar a un interés sustancial hacia esos registros y hacia un aspecto poco conocido de la historia de Australia. El trabajo en la colección fue posible gracias a la profundidad de conocimientos especializados de Bagnall y a las habilidades técnicas y el ingenio de Sherratt. Es importante subrayar que al hacer que las colecciones digitales estén disponibles en la red, las instituciones dedicadas al patrimonio cultural abren las posibilidades para este tipo de uso y reuso. Dada la importancia de estas opciones, valdría la pena que las instituciones las busquen y apoyen más explícitamente, en lugar de sólo esperar a que los usuarios potenciales las encuentren por su cuenta. Para este fin, es preciso considerar cualquier forma en que las instituciones puedan estimular este tipo de usos: a través de concursos, residencias para artistas e historiadores digitales o hackatones en los que los usuarios potenciales puedan colaborar con expertos en el tema y con desarrolladores que trabajen en esas organizaciones.

Cuando vale la pena sacar las pinzas: la remediación de los libros electrónicos del dominio público

Hasta ahora, la mayor parte de los ejemplos pertenecen a la categoría de “mayoreo” como parte del trabajo de las instituciones dedicadas al patrimonio cultural. Los últimos dos ejemplos son de la categoría “boutique”. Son situaciones donde las instituciones del patrimonio cultural se sumergen más en los detalles con el objetivo de alcanzar la remediación o agregar valor al contenido digital para satisfacer las necesidades clave de los usuarios.

Existen millones de libros digitalizados que son del dominio público. Usted puede encontrarlos en una variedad de fuentes y formatos. También puede navegar y buscar en Google Books y Google Ngram o puede leer las obras digitalizadas en la red en HathiTrust y en el Internet Archive. Muchos investigadores utilizan estas herramientas, pero no suele ser lo que la gente elige cuando quiere una lectura divertida o relajante. Entonces, si alguien quiere leer un poco de *Frankenstein* en formato digital cada noche antes de dormir, no van a hacerlo con una copia escaneada que dificulta su lectura. Si quieren disfrutarlo a través de un lector electrónico, es probable que vayan a buscar un archivo EPUB bien formado. La ventaja adicional de ese formato es que también es mucho más accesible para las personas ciegas o con discapacidades visuales. El problema es que, en términos funcionales, esto implica muchas más cosas que simplemente digitalizar el libro.

Se puede encontrar un EPUB para una gran cantidad de libros clásicos en el Project Gutenberg, una red de voluntarios que ha estado recopilando y



publicando copias de textos digitales del dominio público desde 1971. Sin embargo, puede ser complicado navegar esas copias y trasladarlas a un dispositivo de lectura. Alguien debe obtener una copia correcta de la obra en su totalidad para que sea formateada y parezca atractiva para los usuarios. Incluso necesita una portada. El libro en esencia tiene que ser republicado, con mucho del trabajo que conlleva el proceso tradicional de publicación.

Además, existe una variedad de actividades curatoriales que permiten a los usuarios encontrar lo que desean. Si un lector accede a un libro por placer, no necesita ver cientos de ediciones y variaciones de la obra. La facilidad de navegación a través de diferentes dispositivos de lectura es de vital importancia.

Atender las necesidades de un lector así requiere de toda una red de colaboradores. La New York Public Library y la Digital Public Library of America están trabajando en un proyecto llamado Library Simplified. A partir de éste, se ha lanzado una aplicación disponible de forma gratuita que, en muchos casos, proporciona el acceso a obras del dominio público que han sido republicadas, junto con libros electrónicos que se dan en préstamo dentro de la misma plataforma. En relación con el contenido, esto ha significado el hallazgo de copias de calidad particularmente alta de obras del dominio público. Por ejemplo, Standard Ebooks es un esfuerzo de voluntarios que con mucho cuidado republica materiales y pone atención a las ilustraciones, a la corrección de estilo, al marcaje y a la codificación. Al mismo tiempo, el proyecto Library Simplified desarrolló herramientas computacionales para asistir en la organización de datos de cerca de 50 mil títulos del Project Gutenberg para duplicarlos y generar portadas. En el futuro, el uso continuo de estos libros va a depender cada vez más del compromiso de diversas iniciativas para obtener conjuntos de obras de alta calidad que puedan ser agrupados, compartidos y entregados en las formas más deseadas por los usuarios.

El valor agregado: la enseñanza con fuentes primarias

Las colecciones del patrimonio cultural tienen un amplio rango de usuarios potenciales. En muchos casos son investigadores. Gran parte de las convenciones para la descripción, la organización y el acceso se diseñan según las necesidades imaginadas de éstos. Como se demostró con el último ejemplo, hacer el contenido más provechoso y utilizable para audiencias más amplias puede derivar en la necesidad de invertir mayores recursos. Otro grupo de interés para las instituciones del patrimonio cultural es el de los profesores de niveles de educación básicos. De hecho, las colecciones de esas organizaciones están llenas de materiales que pueden ser utilizados por un rango de maestros



que trabajan en campos en los que se usan fuentes primarias de información. Las fuentes históricas son las más obvias, pero existen recursos que son relevantes para el estudio de los idiomas, la literatura y las ciencias. Dicho esto, se requiere mucho trabajo para dar utilidad a estos recursos y que puedan ser aprovechados por los docentes.

Un ejemplo de este tipo de trabajos es el que hace la Library of Congress a través del proyecto Teaching with Primary Sources.¹²³ En él se hace una curaduría de conjuntos de recursos para los profesores de educación básica enfocados en temas que se abordan ampliamente en el programa educativo. Cada uno de estos conjuntos viene acompañado de ensayos breves que contextualizan los elementos y ofrecen sugerencias sobre cómo pueden ser utilizados en las clases. Además, cada conjunto incluye vínculos en los que se pueden encontrar los materiales, así como copias de las reproducciones de las obras en PDF que son fáciles de imprimir en papel de tamaño estándar para que los profesores los usen en clase. Todos los materiales que están agrupados en esta forma se han desarrollado a partir del diálogo con profesores activos para asegurarse de que son adecuados a sus necesidades actuales en el aula. El programa también organiza talleres de desarrollo profesional para que los maestros aprendan la mejor manera de enseñar con este tipo de fuentes y proporciona una variedad de otros materiales relacionados al desarrollo profesional.

Estos son puntos generales para quienes desean que sus colecciones sean utilizadas por los profesores de educación básica. Sin embargo, también demuestran que identificar grupos de usuarios particulares y entender sus necesidades cuando se involucran con el contenido son tareas que pueden tener una inversión de tiempo mucho más prolongada al trabajar con cantidades pequeñas de elementos. Para los profesores, una gran cantidad de objetos es menos útil que una colección temática de recursos específicamente seleccionados y con ideas sobre cómo usarlos para explorar conceptos particulares. Al final, es probable que la vinculación y la continua colaboración con representantes de comunidades de usuarios específicos ofrezcan ideas similares sobre qué tipo de trabajo se podría realizar para satisfacer sus necesidades. Aquí hace falta enfatizar que las colecciones del patrimonio cultural no son un fin en sí mismo. Son herramientas para apoyar el logro de una serie de bienes públicos e institucionales. Entre más tiempo podamos dedicar a averiguar qué tipo de impacto pueden tener nuestras colecciones en las cosas que le importan a la

123 Para conocer el tipo de recursos que desarrolla este programa, ver <http://loc.gov/teachers/>.



gente, más fácil será para todos articular y defender el impacto y los resultados que las colecciones pueden tener en los temas que a las personas les preocupan.

El acceso y las mejorías progresivas

El propósito de la preservación es habilitar el acceso. En muchos casos, éste es una necesidad inmediata. Como los ejemplos en este capítulo han demostrado, el menú de opciones y consideraciones respecto a cómo proporcionar el acceso va siempre en aumento y se vuelve cada vez más complejo. Sin embargo, hay algunos principios y puntos generales por dónde comenzar.

En primer lugar, ser capaz de proveer el acceso al contenido debe ser una prioridad. Usted no debe pasar mucho tiempo pensando en todas las formas posibles de hacer algo accesible antes de simplemente empezar a proporcionar el acceso. La perfección es enemiga de lo bueno, así que el acceso no debe esperar, cuando menos bajo ciertos términos o métodos restringidos. Si usted tiene mucho contenido al que no se puede acceder o no debe proporcionar acceso a él, vale la pena volver a cuestionar si en verdad debería tenerlo. Si no se va a usar, ¿para qué sirve entonces?

En segundo lugar, comprender a los individuos que figuran en las colecciones es un requerimiento de primer orden. ¿Cuánto tiempo tienen las colecciones? ¿Todavía están vivos los individuos que aparecen en ellas? ¿Existen consideraciones en cuanto a los derechos de autor? ¿Qué hay de la privacidad? Usted debe asegurarse de siempre pensar en las vidas de las personas que están entrelazadas con cualquier colección con la que trabaje y de que se entienden los problemas legales que rodean el acceso a un cierto conjunto de materiales. Estos puntos evidencian el tipo de restricciones que es necesario implementar. Por lo general, el primer paso es verificar que no se perjudica a nadie. Sin embargo, de acuerdo con los valores de los archivos y las bibliotecas profesionales, es necesario poner especial atención al daño que las personas de los grupos oprimidos podrían sufrir, ya que se trata de quienes han sido repetidamente explotados por las instituciones del patrimonio cultural y son los más propensos a verse afectados si la información sobre ellos es mal utilizada. A partir de esto, es fundamental pensar en la variedad de métodos que se pueden usar para restringir el acceso. Puede tener sentido descartar y eliminar el contenido que tenga un valor limitado, pero que podría provocar un daño significativo. Es posible que sea preferible restringir el acceso a una sala de lectura o a aquellos con una cuenta institucional.

En tercer lugar, hay que estar consciente de que el trabajo para proveer el acceso en el presente sólo es el comienzo y no el final. Usted puede subir



algunos conjuntos de datos masivos con mucha rapidez y más adelante puede regresar a ellos con el objetivo de producir interfaces para los materiales a nivel de los elementos. Puede proporcionar el acceso a los elementos con registros mínimos y, después, cuando lo permita el tiempo, ofrecerlos mejorados y más detallados. Debe establecer planes para ver cómo son utilizadas las colecciones y explorar posibles alianzas con los usuarios para que puedan crear maneras de tener acceso a las colecciones o, como lo demuestra el caso del Internet Archive con GifCities, puede tener una forma estándar a largo plazo para proporcionar el acceso al contenido (como la Wayback Machine) y aún así extraer algún subconjunto especial del contenido y proveer una interfaz experimental adicional.

En cuarto lugar, algunas de las mejores y más atractivas interfaces y modos de acceso a las colecciones pueden ser desarrolladas por los usuarios. Es posible que esto sea evidente, como en los casos de *Invisible Australians*, *One Terabyte of the Kilobyte Age* y *Viral Texts*, en los que cada vez más los resultados del uso académico y artístico de las colecciones son en sí mismos interfaces y medios de acceso a ellas. Lo anterior abre un área de trabajo aún emergente y muy importante. ¿Cómo pueden las bibliotecas, los archivos y los museos facilitar la producción de este tipo de interacciones con las colecciones? Más allá de eso, ¿cuáles son los roles que pueden asumir esas instituciones para ayudar a administrar, recopilar, preservar y proporcionar el acceso a este tipo de conjuntos de datos derivados y puntos de acceso a los contenidos? Ya que los usuarios son quienes están creando las interfaces más útiles para las colecciones, las instituciones deben tener un claro interés en ayudar a encontrar maneras de asegurar que sean incorporadas y preservadas en el futuro.

En quinto lugar, uno de los modos más poderosos para permitir el acceso proviene del desarrollo de conjuntos de datos derivados. Como varios ejemplos lo han demostrado, proporcionar el acceso masivo desemboca en el desarrollo de medios muy creativos de interacción y exploración de las colecciones que son creados por académicos y artistas. Más allá de esos modos de acceso masivo, en muchos casos, algunas de las formas más útiles para los datos han sido las derivadas. Aunque no se puede usar el *software* de la NSRL, los valores de *hash* publicados son importantes para una variedad de usuarios y son un medio para identificar de manera única los archivos de *software*. Asimismo, a pesar de que no se puede compartir toda la información de Google Books, debido a los derechos de autor, los n-gramas y el visor Ngram se han convertido muy rápido en una manera poderosa de estudiar algunas tendencias a través del tiempo.



En sexto lugar, hay que entender que algunas de las maneras más sustanciales de hacer que el contenido sea accesible para un público en particular pueden implicar un trabajo extenso dedicado a pequeños números de objetos que pueden ser más útiles para esos usuarios. Un millón de libros digitalizados son geniales para un académico, pero cuando uno desea relajarse y leer algo de literatura antes de dormir tal vez una lista bien curada de EPUB exquisitamente reformateados y publicados sería mucho más provechosa. De la misma forma, millones de imágenes históricas digitalizadas podrían ser buenas para un investigador profesional, pero un subconjunto de objetos curados con instrucciones e ideas sobre cómo utilizarlos cumpliría mejor con las necesidades de un profesor de educación básica. Existen ocasiones en las que los enfoques especializados en una cantidad pequeña de subconjuntos de materiales pueden traer beneficios significativos, pues satisfacen las necesidades de los grupos de usuarios específicos, y es bueno poder identificarlos.

Las posibilidades del acceso digital son emocionantes y emergentes. La buena noticia de todo esto es que hay cosas muy simples que podemos hacer desde el comienzo. Después debemos trabajar en conectar cada vez más con varios grupos de usuarios para descubrir cuál es la mejor manera de invertir los recursos limitados que tenemos para proporcionar el acceso a los diversos tipos de contenido. Conforme obtengamos datos derivados de las colecciones, al tiempo que otros puedan hacer uso de ellos y produzcan variantes y otro tipo de copias, notaremos que las características fundamentales de la tradición de la preservación han sido esenciales para asegurar el acceso a largo plazo. En términos generales, se piensa que conservar artefactos y hacer copias perfectas de la información es el trabajo profesional de la preservación, pero, en realidad, son los métodos folclóricos, o vernáculos, de compartir y volver a contar y reformatear y copiar, que han sido las fuerzas más antiguas de esta labor en la historia humana. Es importante que no olvidemos esto cuando permitimos que el mundo utilice, reutilice y haga mezclas de nuestras colecciones.

Conclusiones

Herramientas hacia el futuro

La preservación digital no es una ciencia exacta. Es un oficio en el que los expertos deben usar y refinar conscientemente sus criterios para evaluar el contenido digital e implementar estrategias que minimicen el riesgo de pérdida. Al menos eso es lo que he intentado argumentar en este libro.

El oficio de la preservación digital tiene sus raíces en el pasado. Se construyó a partir de los registros, los archivos y la labor de quienes nos precedieron y de aquellos que diseñaron y configuraron los sistemas que permiten la creación, la transmisión y la representación de su trabajo. Al mismo tiempo, el oficio de la preservación digital es también el de un futurista. Debemos reconocer las tendencias del pasado en el siempre cambiante desarrollo de los medios digitales y hacer nuestra apuesta sobre cómo progresarán las tecnologías digitales en el futuro.

Uno de los mayores retos de escribir cosas prácticas sobre las herramientas y tecnologías digitales es lo rápido que los cambios tecnológicos pueden hacerlas parecer anticuadas. Espero que al estar enfocado en las maneras de abordar y tomar decisiones sobre los objetos, este libro tenga una larga vida en los estantes. Sin embargo, entiendo que es posible que se den cambios significativos en las tecnologías digitales básicas que puedan tener resultados muy diferentes en ciertas situaciones en el futuro. Por eso, voy a cerrar el libro con la discusión y evaluación de una serie de tecnologías, tendencias y movimientos sociotécnicos relevantes para la planeación de la preservación digital. Mi esperanza es que esto conduzca a una reflexión sobre algunas tendencias y cambios tecnológicos a corto plazo que podrían ser relevantes para la preservación digital. En un sentido más amplio, quisiera mostrar herramientas que ayuden a pensar más a fondo sobre el impacto potencial de cualquier tendencia.



Las tecnologías emergentes y los desvíos tecnológicos

Existe una serie de tecnologías emergentes y tendencias tecnológicas que son importantes para pensar el futuro de la preservación digital. Analizaré brevemente dos de ellas: los paradigmas de las interfaces de usuario emergentes y los cambios en la tasa de crecimiento de la capacidad de almacenamiento. Tras discutir estas dos tendencias, revisaré otra variedad que yo clasifico como desviaciones. Los detalles de cada una de estas tecnologías son potencialmente relevantes en sí, pero es aún más significativo el proceso mental utilizado para considerarlas, ya que ilustra cómo se deben abordar y evaluar las señales y tendencias de las tecnologías como profesional de la preservación digital. No tome esta sección del libro como un conjunto de predicciones, sino, más bien, como un ejemplo del ejercicio de predicción en el que, como profesional de la preservación digital, se debe involucrar.

Después de evaluar las tendencias tecnológicas emergentes a corto plazo, ampliaré la perspectiva y haré algunas observaciones más generales sobre los desafíos que implica trazar el futuro de las tecnologías digitales para trabajar en instituciones del patrimonio cultural.

Los cambios en la lógica de la interfaz

Desde su lanzamiento masivo en el mercado con el iPhone, las interfaces táctiles se han convertido en la vía principal con la que los usuarios alrededor del mundo interactúan con las computadoras. Esta transición es una excelente oportunidad para revisar nuestras suposiciones acerca del uso futuro de los contenidos digitales. Si esta tendencia continúa, podemos llegar al punto en el que cosas como los teclados simplemente dejen de formar parte de la configuración predeterminada de un ambiente de cómputo. En ese caso, varios contenidos que asumían su uso necesitarían ser adaptados para poder utilizarse. Muchos dispositivos tipo tabletas contienen teclados virtuales, lo que nos da idea de lo que podría suceder: cada vez más capas de interfaces virtuales en las tecnologías.

Los cambios que ha traído la computación que se basa en lo táctil pueden ser los primeros de una serie de novedosos paradigmas de la interfaz. Los dispositivos de realidad virtual, como Oculus y HTC Vive, le han dado vida a lo que por décadas fue sólo producto de la imaginación: una tecnología de consumo factible. ¿Qué oportunidades habrá para proporcionar el acceso al contenido en un ambiente de realidad virtual? Y al mismo tiempo, ¿qué tipo de consideraciones éticas entrarán en juego al colocar a alguien en una experiencia más encarnada y directa con el contenido?



De alguna manera, el mayor desafío conceptual de las nuevas modalidades de las interfaces de los ambientes informáticos es la voz. Herramientas como Alexa de Amazon y Google Echo están empezando a aparecer en casas de todo el país y ofrecen una interfaz para sistemas informáticos completamente no visual. ¿Qué implicaría hacer que las colecciones digitales fueran utilizables y accesibles a través de interfaces basadas en la voz? Cada una de ellas sería tanto el contexto que debemos considerar para recopilar y preservar formas novedosas de contenido digital original como el contexto al cual necesitaríamos transferir el contenido digital histórico para hacerlo accesible y utilizable. El caso de las tecnologías de voz plantea importantes cuestionamientos en relación con los sistemas que están “siempre encendidos” y graban todo el sonido a su alrededor y con la manera en que recopilan y administran esa información. A la par, el contenido en estos sistemas invita a la reconsideración de los factores de sustentabilidad de los formatos digitales. Si resulta que se trata de contenido fundamental vinculado sólo a las características de alguna de estas nuevas plataformas de interfaz y a su naturaleza cada vez más propietaria, el asunto se vuelve problemático. Estas lógicas variables de las interfaces presentan una serie de desafíos, pero también existen algunas tendencias en el desarrollo del almacenamiento digital básico que plantean sus propios retos.

La tasa de Kryder se ha ido desacelerando

En los últimos 30 años nos hemos acostumbrado a que en poco tiempo deberíamos poder comprar nuevos volúmenes de almacenamiento digital significativamente mayores por un precio cada vez menor. Nuestra experiencia como consumidores ha sido que si compramos una computadora nueva cada cinco o seis años, ésta tendría un incremento dramático en su capacidad de almacenamiento. Recuerdo con claridad que en algún momento adquirí una computadora con ocho *gigabytes* de almacenamiento y pensé: “¿Cómo podría llegar a llenarla?”. Desde luego que lo hice, todos lo hicimos, ya que con más espacio de almacenamiento se hicieron posibles más y más tipos de archivos más grandes. El ritmo de la caída casi exponencial de los precios de almacenamiento se explica con la Ley de Kryder. Nombrada en honor del director técnico de la empresa Seagate, el concepto se enfoca en la idea de que uno podría esperar que los discos guarden 40% más de datos en el mismo espacio físico cada año como producto de los avances en la tecnología de almacenamiento.¹²⁴

124 Chip Walter, “Kryder’s Law”, *Scientific American* (2005): 32-33.



Esta tasa de incremento se ha estancado durante los últimos años. Lo que principalmente se desprende de este cambio es que “el almacenamiento será mucho menos libre de lo que solía ser”.¹²⁵ Es importante destacar que el ritmo de crecimiento de la creación de contenidos digitales no ha disminuido. Asimismo, es probable que el almacenamiento consuma cada vez más recursos y presupuesto de las instituciones que se dedican a las tareas de la preservación digital.¹²⁶ Con esto en mente, si la tasa de Kryder continúa estable, existe una alta probabilidad de que afecte el tipo de contenidos que se producen y que abra la posibilidad de desarrollar medios de almacenamiento digital más longevos o con arquitecturas diferentes,¹²⁷ un tema al que volveré más adelante.

La conclusión de esta tendencia es que los costos del almacenamiento comenzarán a aumentar para las instituciones del patrimonio cultural que se toman en serio la preservación digital. El impacto de esto no está del todo claro, pero parece que llevará a estas organizaciones a ser más selectivas en lo que resguardan y en lo que se comprometan a preservar a largo plazo, al tiempo que las grandes instituciones continuarán invirtiendo en la investigación de sistemas de almacenamiento jerárquico que utilicen distintos medios de almacenamiento para mantener los costos bajos.

La discusión sobre estas dos tendencias tecnológicas en particular (las nuevas interfaces y los cambios en la trayectoria de los medios de almacenamiento) nos ofrece modelos para identificar los potenciales impactos que otras tendencias del panorama tecnológico general pudieran tener en nuestra labor. Dicho esto, quiero dedicar un poco de tiempo a hablar sobre algunas de las trampas que pueden surgir si miramos demasiado hacia el futuro en busca de respuestas para los problemas de hoy.

La trampa del solucionismo tecnológico

Jamás habrá una aplicación para la preservación digital. En relación con este punto, volvemos al deseo de tener un supersistema único que “simplemente

125 David Rosenthal, “Storage Will Be a Lot Less Free than It Used to Be”, *DSHR's Blog*, 1o. de octubre de 2012, <https://blog.dshr.org/2012/10/storage-will-be-lot-less-free-than-it.html>.

126 Rosenthal *et al.*, “The Economics of Long-Term Digital Storage” (ponencia, “The Memory of the World in the Digital Age: Digitization and Preservation UNESCO Conference”, 26-28 de septiembre de 2012).

127 Ian F. Adams, Ethan L. Miller y David Rosenthal, “Using Storage Class Memory for Archives with DAWN, a Durable Array of Wimpy Nodes”, *Technical Report UCSC-SSRC-11-07* (Santa Cruz, California: Storage Systems Research Center, University of California, Santa Cruz, 2011).



resuelva” la preservación digital del que hablamos en la introducción. De la misma manera, no habrá una aplicación que solucione los desafíos más apremiantes de la sociedad, como la justicia social, la equidad económica, los derechos civiles, la pobreza o un mejor medio ambiente. Sin embargo, la rapidez con la que la informática ha cambiado nuestras vidas a veces nos hace pensar que el día de mañana puede salir una aplicación capaz de “resolver” los eternos problemas sociales, políticos y económicos. No hay que apostar por que la vinculación de información, los *blockchains*, los códigos QR, las *bitcoin* o la inteligencia artificial puedan ofrecer o sugerir nuevas maneras de solucionar nuestros problemas. Tampoco hay que esperar a la liberación del nuevo sistema para repositorios de código abierto para empezar a hacer la preservación digital. La tecnología no nos salvará de la necesidad de levantar las cajas de información del piso antes de la inundación, metáfora de la que hablé antes.

De acuerdo con gran parte del discurso de la informática en la sociedad contemporánea, existe el impulso de “resolver” el problema de la preservación digital a través del solucionismo tecnológico. Yo he sugerido que no se trata sólo de un problema, sino de innumerables dificultades locales que dependen de lo que cada comunidad considera valioso.

Uno de los grandes problemas de la preservación digital es que muchos creen con mucha firmeza que todo el asunto radica en el aspecto técnico o que la preservación digital es un problema que puede resolverse. He tomado prestado el término “solucionismo tecnológico” de Evgeny Morozov que, a su vez, lo tomó prestado del concepto “solucionismo” usado en la arquitectura. Este crítico explica que “el solucionismo presupone que en lugar de investigar el problema que se está tratando de resolver, se busque la respuesta antes de que las preguntas se hayan terminado de formular por completo”.¹²⁸ Dicho de otro modo, la preservación digital, es decir, el asegurar el acceso a la información digital a largo plazo, no es necesariamente el simple problema de mantener el contenido digital disponible, sino un complejo y multifacético dilema sobre qué es lo más importante dentro de todo este contenido digital en distintos contextos, tanto actuales como futuros.

El solucionismo tecnológico de la informática en la sociedad contemporánea puede seducirnos fácilmente y hacernos pensar de manera errónea que existe una especie de “botón de conservar” o que uno puede dar clic derecho al folder de “Cultura americana” en el metafórico escritorio del mundo

128 Evgeny Morozov, *To Save Everything, Click Here. The Folly of Technological Solutionism* (Nueva York: Public Affairs, 2013), 6.



y seleccionar la opción “conservar como...”. Mi esperanza es que este libro haya mostrado el peligro que conlleva esta ilusión-tecno-fantástica. En última instancia, el futuro del acceso a la información digital es, como ha sido el caso a lo largo de la historia, una cuestión de recursos. ¿A quién le daremos los recursos para tomar decisiones respecto al futuro de nuestro pasado? ¿Valoramos nuestras instituciones de la memoria cultural lo suficiente como para apoyarlas y mantenerlas en esta tarea? ¿Hasta qué punto administramos nuestro mundo natural y nuestros recursos de una manera que nos permita que la futura preservación, incluso de nuestra propia especie, sea viable? Me centraré más en estos cuestionamientos en el resto del capítulo, pero antes quiero profundizar un poco más en algunas cuestiones clave sobre la predicción del futuro de las tecnologías digitales.

Hay que recordar quién tiene el control

Lo que yo quiero para el futuro de la tecnología digital es muy irrelevante. Nuestra sociedad provee y abastece de recursos a la imaginación de nuestro futuro tecnológico y no a nuestras comunidades, maestros, bibliotecarios, archivistas, historiadores, sociólogos o científicos. Todos nosotros trabajamos bajo la sombra proyectada de dos personajes que sí tienen el poder para soñar un futuro y manifestarlo como realidad a través de la aplicación de recursos.

La industria militar y Silicon Valley son quienes en realidad pueden imaginar el futuro y tienen los recursos necesarios para manifestar su visión de esos mundos posibles y tecnológicos en la realidad. La red, la minería de datos, el reconocimiento facial, la criptografía, los sistemas de almacenamiento de datos, todos estos sistemas e infraestructuras nos llegan desde sus mentes, y el enorme flujo de capital de inversión que llega a Silicon Valley y del presupuesto del gobierno federal de Estados Unidos destinado a la investigación y al desarrollo militar. Así es como se moldea el futuro en este país y el resto del mundo. Si no se llevan a cabo cambios radicales en la manera en que se asignan los recursos a nivel mundial, cosa que parece bastante improbable, esto no va a cambiar.

El resultado de esto es que los archivistas, los historiadores, los humanistas y todos los otros sectores de la sociedad civil no reciben los recursos necesarios para dirigir ninguna de las decisiones fundamentales sobre qué tipo de tecnología vale la pena hacer realidad en el futuro. Es un panorama un poco sombrío, pero creo que es mejor ver las cosas tal como son, ya que esto tiene algunas implicaciones pragmáticas en nuestro trabajo. Nuestra misión en el orden social existente es imaginar cómo apropiarnos de esas herramientas creadas para vigilar y controlar y encontrar maneras de usarlas para proteger



y preservar. Es esencial que no hagamos este trabajo a ciegas. Las herramientas de la tecnología digital no están aquí para emanciparnos. Son herramientas creadas por las estructuras de poder de nuestra sociedad con la intención de construir su propia visión del futuro.

Es importante recordar que las nuevas tecnologías no surgen de la nada. Las computadoras de hace 20 años no son inherentemente obsoletas. No avanzamos en línea recta hacia un futuro más eficiente en términos objetivos. Cada eficiencia asume sus propias prioridades. La gestión de los derechos digitales es una especie de “gen exterminador”, no muy distinto a Monsanto y sus semillas que no se pueden reproducir, pues construye tecnologías bajo el supuesto de la obsolescencia programada y crea innovaciones disruptivas, acciones vinculadas a una batalla continua que ha sido descrita por muchos como “capitalismo tardío”.

El problema de las instituciones culturales en la era del neoliberalismo

“¿Necesitamos bibliotecas en la era de Google?”. Éste es el tipo de declaraciones absurdas que sólo son posibles en una sociedad seducida por el canto de sirena del capitalismo tardío. ¿Cómo puede ser culturalmente posible que se formulen este tipo de preguntas? La idea de que una compañía sin fines de lucro, fundada hace menos de 20 años, pueda llevar a la gente a cuestionar la función social, cívica y económica que brindan las más de cien mil bibliotecas, tan sólo en Estados Unidos, ejemplifica una creciente enfermedad en nuestra manera de pensar. No, los ingresos provistos por la publicidad en los buscadores no van a crear una infraestructura nacional para la recopilación y la preservación que proporcione el acceso al conocimiento y la sabiduría y que apoye el aprendizaje continuo.

No pretendo culpar a Google. El problema reside tanto en nuestra falta de confianza en las instituciones como en la fe ciega en que el éxito del desarrollo de la tecnología informática puede llevar a una transformación completa de la sociedad. Las ideas de innovación disruptiva de Silicon Valley ocupan mucho espacio en nuestra imaginación cultural porque, en buena medida, se ajustan ergonómicamente al antiguo impulso social de deshacerse de los bienes sociales, cívicos y públicos.¹²⁹ Esta utopía tecnológica se presenta al mismo tiempo que experimentamos la larga historia del neoliberalismo.

129 Siva Vaidhyathan, *The Googlization of Everything* (Berkeley: University of California Press, 2012).



Bajo la lógica del neoliberalismo casi todos los aspectos de la sociedad están imbuidos en la lógica del mercado. Los Estados disponen de cada vez menos recursos para las bibliotecas y los archivos, mientras que algunas empresas se han inclinado por la digitalización de sus colecciones, siempre y cuando se les conceda por tiempo limitado el monopolio de digitalizar el contenido que ellos mismos elijan y que le puedan dar acceso sólo a las personas que paguen suscripciones.¹³⁰ Una de las mayores amenazas para la sostenibilidad de la información digital, actual y futura, es la total o parcial falta de financiamiento de nuestras instituciones culturales y sociales. En este sentido, una de las mayores cosas que podemos hacer para apoyar la preservación digital es demostrar el valor y la relevancia del trabajo que hacemos a las comunidades que servimos.¹³¹ No podemos dar por hecho que la sociedad continuará creyendo que las instituciones del patrimonio cultural son bienes públicos que vale la pena mantener y que es necesario desarrollar las capacidades de nuestras organizaciones para probar que son valiosas.

También vale la pena hacer hincapié en lo insostenible e inequitativa que se ha vuelto la labor de los trabajadores dedicados al patrimonio cultural. A medida que los recursos son más escasos, los proyectos se agotan y hay menos puestos disponibles. Estos empleos pagan cada vez menos a los entusiasmados y comprometidos jóvenes archivistas, bibliotecarios y profesionales de los museos. Lo anterior es particularmente problemático porque estas áreas de trabajo de manera tradicional han sido ocupadas en su mayoría por mujeres y porque bajo las normas patriarcales de nuestra cultura estos roles han sido mal pagados y subestimados a lo largo de la historia. El asunto se vuelve aún más complejo al adentrarnos en la *gig economy*,¹³² soñada y hecha realidad por compañías como Uber, Airbnb y TaskRabbit. Uno de los ejemplos más preocupantes es que algunas instituciones dedicadas al patrimonio cultural han empezado a contratar a los sistemas penitenciarios para hacer su trabajo. Aunque la idea de dar oportunidad a los presos de aprender a manejar el patrimonio cultural suena potencialmente prometedora como un camino de desarrollo personal significativo y una oportunidad laboral, el hecho de que esas carreras requie-

130 Adam M. Kriesberg, "The Changing Landscape of Digital Access: Public-Private Partnerships in US State and Territorial Archives" (tesis de doctorado, University of Michigan, 2015).

131 Para una excelente explicación sobre el concepto de relevancia en las instituciones culturales, ver Nina Simon, *The Art of Relevance* (Santa Cruz, California: Museum 2.0, 2016).

132 N. del T: el término en inglés *gig economy* se refiere a un modelo de trabajo basado en pequeños encargos o bajo demanda.



ran posgrados avanzados ha provocado que esta tendencia termine siendo una fórmula neoliberalista perfecta para el desastre. En lugar de encontrar el modo para que la gente pueda tener carreras bien pagadas en bibliotecas, archivos o museos, nuestra sociedad se las ha ingeniado para que tanto las comunidades como las prisiones proporcionen mano de obra barata, con lo que se mina aún más el futuro de una fuerza de trabajo profesional del patrimonio cultural.

Los problemas del neoliberalismo y la *gig economy* han dificultado aun más que la fuerza laboral del patrimonio cultural pueda crecer para reflejar mejor la diversidad de nuestras comunidades. Existe un claro y explícito deseo de que los equipos de trabajo de las instituciones dedicadas a ese ámbito muestren esa pluralidad de las comunidades a las que sirven. Sin embargo, dada la aparente necesidad de que los recién egresados, a menudo con múltiples títulos, pasen a través de una serie de pasantías sin goce de sueldo y compitan a nivel nacional por contratos temporales mal pagados para entrar a la arena de esta profesión, estamos en una situación en la que sólo los más privilegiados pueden darse el lujo de trabajar en el área del patrimonio cultural. Así, nos enfrentamos con dificultades que sólo podremos resolver si nos abocamos a establecer mejor la fuente de recursos de nuestras instituciones y a integrar mejor nuestra fuerza laboral en las comunidades para las que trabajamos.

Debo enfatizar que no creo que el estado actual de lo que he descrito en realidad sea un sistema social intencionalmente construido que está haciendo las cosas para las que fue diseñado. En vez de eso, lo veo como el enredo de una serie de marcos ideológicos que no son sofisticados, en particular, ni están bien desarrollados. El panorama es desolador, pero creo que es mejor ser honestos con nosotros mismos y no engañarnos. Los que hemos tenido éxito en las instituciones de educación superior y cada vez más en las bibliotecas y los archivos no debemos convencernos de que ha sido porque hemos trabajado duro. Estoy seguro de que lo hemos hecho, pero no es suficiente. Es responsabilidad de quienes ya nos encontramos en los botes salvavidas de una carrera dentro de las instituciones dedicadas al patrimonio cultural pensar en qué podemos y qué debemos hacer para ayudar a aquellos que quieren entrar. Siempre habrá presiones, y nosotros tenemos el deber de hacer lo que podamos para que los sistemas cada vez más opresivos en los que funcionamos sean lo más humanos posibles. Tenemos la obligación y la responsabilidad de que en nuestras instituciones y campos de trabajo prevalezca una ética del cuidado.¹³³ Debemos

133 Bethany Nowviskie, “On Capacity and Care” (discurso, “National Endowment for the Humanities-Digital Humanities Project Directors’ Meeting”, septiembre de 2015).



nadar contra corriente y crear espacios laborales con pagos adecuados y tiempo libre con los que se pueda vivir una vida plena. La corriente, por decirle de alguna manera, es una metáfora importante para nuestra siguiente consideración. Esta desolación que ha traído el neoliberalismo es manejable en comparación de los desafíos que nos ha llevado a enfrentar con nuestro entorno natural.

Las instituciones dedicadas al patrimonio cultural en el Antropoceno

El siglo XX trajo consigo algo más novedoso que la aparición de los medios digitales: la habilidad de alterar de manera radical nuestro mundo al punto de volverlo inhóspito para la humanidad. Hemos entrado al Antropoceno. No está del todo claro si va a ser una era, una época o un evento.¹³⁴ ¿Reconoceremos el poder de la tecnología y la ciencia y nos volveremos mejores administradores de nuestro pálido punto azul o continuaremos descuidadamente rumbo a nuestra posible extinción? Sólo el tiempo lo dirá, pero el panorama general no se ve nada bien.

El cambio climático mundial antropogénico está sucediendo. La ciencia está segura. En la próxima mitad del siglo veremos dramáticas alteraciones del medio ambiente en el mundo entero y el resultado de esto tendrá un impacto radical en todos los sectores de la sociedad, incluyendo las instituciones dedicadas al patrimonio cultural. Para contextualizar, sólo en Estados Unidos, más de la mitad de las ciudades más importantes están a menos de diez pies arriba del nivel del mar.¹³⁵ Muchas de las instituciones dedicadas al patrimonio estarán literalmente bajo el agua el próximo siglo. Éste es un tema al que los bibliotecarios, los archivistas y los profesionales de los museos están respondiendo con una actitud proactiva, a través de iniciativas como *Archivists Respond to Climate Change* y *Keeping History Above Water*.¹³⁶ Vale la pena señalar que el National Parks Service también ha proporcionado una guía sobre la planificación de escenarios para los sitios históricos, una buena herramienta para

134 David Grinspoon, *Earth in Human Hands. Shaping Our Planet's Future* (Nueva York: Grand Central, 2016).

135 Benjamin Strauss, "What Does U. S. Look Like with 10 Feet of Sea Level Rise?", *Climate Central*, 13 de mayo de 2014, <http://www.climatecentral.org/news/us-with-10-feet-of-sea-level-rise-17428>.

136 Para más sobre estas iniciativas y oportunidades para participar, ver Eira Tansey, "Archival Adaptation to Climate Change", *Sustainability: Science, Practice & Policy* 11, núm. 2 (2015).



que cualquier institución dedicada al patrimonio cultural comience a pensar en ejercicios de planeación sobre cómo continuará cumpliendo con su misión al enfrentar un cambio medioambiental.¹³⁷ A lo largo de todo este trabajo, ha quedado claro que lo más probable es que ocurran más desastres naturales en todo el mundo, lo cual hace que sea más crucial que las instituciones dedicadas al patrimonio cultural desarrollen proyectos para saber cómo responder ante catástrofes en sus comunidades y, en términos ideales, cómo ayudar a otras.

En este contexto, es cada vez más importante que las instituciones del patrimonio cultural exploren maneras de ser más sustentables en su relación con el medio ambiente. El repertorio circular de los nuevos dispositivos computacionales siempre más elegantes de la privilegiada minoría del mundo se sostiene gracias a las condiciones laborales muy problemáticas en la mayoría del mundo, en la explotación de una variedad de recursos naturales y en fábricas que acarrear serios problemas ambientales y sociales. Además, no sólo hay que tomar en cuenta el problema de la producción de las tecnologías informáticas, también existe el gran dilema de dónde acaban al ser desechadas cada vez más rápido. Los residuos electrónicos tienen significativos efectos nocivos en la salud del ser humano en países como China e India.¹³⁸ Bajo este panorama, es importante que los encargados de la preservación digital se comprometan a establecer prácticas verdaderamente ecológicas. Esto puede y debe incluir la huella de carbono de los servidores y la manera en la que las instituciones pueden adoptar estrategias de reducción de residuos electrónicos. Cuando se consideran los costos de calefacción y enfriamiento, tanto en términos de huella de carbono como en el costo de la electricidad, se pueden tomar decisiones diferentes y más eficientes en cuanto a los medios de almacenamiento. De hecho, algunas instituciones del patrimonio cultural ya han empezado a explorar estos temas.¹³⁹ A la par de este cambio, si además asumimos que la tasa de Kryder seguirá manteniéndose, podríamos enfocarnos en invertir en sistemas de almacenamiento más duraderos, más eficientes en términos de consumo de energía

137 Matthew Rose y Jonathan Star, *Using Scenarios to Explore Climate Change: A Handbook for Practitioners* ([Washington DC]: National Park Service, U. S. Department of the Interior, 2013).

138 F. O. Ongondo, Ian D. Williams y Tom Cherrett, "How Are WEEE Doing? A Global Review of the Management of Electrical and Electronic Wastes", *Waste Management* 31, núm. 4 (2011): 714-730.

139 Rosenthal *et al.*, "Green Bytes: Sustainable Approaches to Digital Stewardship" (Panel de discusión, "Library of Congress Digital Preservation 2013 Meeting", 23-25 de julio de 2013).



y que produzcan menos desechos electrónicos. En este sentido, los esfuerzos de las humanidades digitales como la informática mínima, que busca maneras de usar tecnologías y sistemas computacionales más sustentables, se han convertido en sitios muy importantes para el futuro de la preservación digital.¹⁴⁰

Los sitios de esperanza en un mundo roto

Quisiera concluir con un tema que me da esperanza. Creo que hemos estado bastante engañados sobre dónde ocurre la innovación en el mundo. Mientras que la riqueza y el poder se han acomodado para que un pequeño grupo de gente privilegiada diseñe nuestro futuro, un elenco muy diferente de personajes ha mantenido nuestro futuro a flote durante mucho tiempo. Stephen Jackson, en su ensayo “Rethinking Repair”, nos invita a “considerar la erosión, el deterioro y la descomposición como nuestros puntos de partida, en lugar de la novedad, el crecimiento y el progreso, al pensar en la naturaleza, el uso y los efectos de las tecnologías de la información y los nuevos medios”.¹⁴¹ Esta línea de pensamiento nos lleva a ver la capacidad que tiene la civilización humana para absorber y responder a los cambios radicales como el punto de partida para la innovación. A través de una discusión sobre la industria del desmantelamiento de buques en Bangladés, Jackson demuestra cómo una fuente mundial de desechos y restos de barcos abandonados se puede convertir en un recurso que se descompone en partes y regresa a la circulación en la economía global.

En este contexto, no debemos enfocarnos en la fuerza continua de la “innovación disruptiva”, ya que es precisamente esto lo que ha llevado al mundo al borde del desastre. En cambio, debemos volcar nuestra atención en las fuerzas del mantenimiento y la reparación que surgen a partir de estos supuestos innovadores. De hecho, ha aparecido una comunidad de estudiantes y profesionales que se han concentrado en la idea de volver a poner al mantenimiento en el centro de nuestra concepción del futuro.¹⁴² Las instituciones del patrimonio cultural pueden beneficiarse de forma considerable de este cambio

140 “Minimal Computing”, Global Outlook: Digital Humanities, acceso el 28 de abril de 2021, www.globaloutlookdh.org/minimal-computing/.

141 Steven J. Jackson, “Rethinking Repair”, *Media Technologies Essays on Communication, Materiality, and Society*, ed. de Tarleton Gillespie, Pablo J. Boczkowski y Kirsten A. Foot (Cambridge, Massachusetts: MIT Press, 2014), 221.

142 Andrew Russell y Lee Vinsel, “Hail the Maintainers”, *Aeon*, 7 de abril de 2017, <https://aeon.co/essays/innovation-is-overvalued-maintenance-often-matters-more>.



de foco en la mentalidad de nuestra cultura. Sin embargo, dicho esto, tenemos casi tanto que ganar si también enfocamos en esa dirección nuestras propias concepciones de los roles y funciones de nuestras instituciones. Con demasiada frecuencia, las ideologías subyacentes a las tecnologías digitales influyen en las ideas sobre el trabajo de la preservación digital. Para entender el futuro de las instituciones del patrimonio cultural debe ser fundamental comprender que la preservación digital, o de cualquier otro tipo, implica comprometer recursos financieros y empoderar a las generaciones actuales y futuras de bibliotecarios profesionales, archivistas y profesionales de los museos para que construyan y mantengan la infraestructura social y cívica del papel que nuestras instituciones tienen al servicio del mundo. No podemos predecir el futuro, pero sí podemos invertir en él.

Nuestro papel en la carrera de relevos

Martha Anderson, ex directora ejecutiva del National Digital Information and Infrastructure Program de la Library of Congress, a quien tuve el honor de conocer como mentora y guía para desarrollarme en este oficio, describió la preservación digital como una carrera de relevos. La preservación digital no se trata de un sistema en especial o de una serie de acciones de preservación. Se trata de preparar contenido y colecciones para la primera de una gran cadena de entregas. No podemos predecir cuáles serán los futuros medios e interfaces digitales ni cómo funcionarán, pero podemos seleccionar los materiales actuales y trabajar con las comunidades para articular con ellas los aspectos que sean importantes para sus casos particulares, hacer copias perfectas y luego esforzarnos para proteger nuestras apuestas en las tendencias de la tecnología digital para tratar de hacer el próximo pase de la manera más fluida posible. Eso es lo que significa practicar el oficio de la preservación digital.

No hay final para la preservación digital. Lo más que podemos esperar es ser un eslabón en una cadena ininterrumpida de memoria. Dados los desafíos que enfrenta nuestro mundo, creo que esos vínculos con nuestro pasado, esas conexiones con los hechos y las decisiones que tomamos sobre qué historias son lo suficientemente importantes como para constituir nuestra memoria social colectiva ahora son más importantes que nunca. Espero que este libro pueda servir de alguna manera como una guía útil para aquellos de ustedes que desean ayudar a mantener y reparar esa cadena.

Bibliografía

- “A Wonderful Invention—Speech Capable of Indefinite Repetition from Automatic Records”. *Scientific American* (17 de noviembre de 1877).
- “About Documentary Editing”. Association for Documentary Editing. Acceso el 28 de abril de 2021. http://www.documentaryediting.org/wordpress/?page_id=482.
- “About the Archimedes Palimpsest”. Archimedes Palimpsest Project. Acceso el 28 de abril de 2021. <http://archimedespalimpsest.org/about/>.
- Adams, Ian F., Ethan L. Miller y David Rosenthal. “Using Storage Class Memory for Archives with DAWN, a Durable Array of Wimpy Nodes”. *Technical Report UCSC-SSRC-11-07*. Santa Cruz, California: Storage Systems Research Center, University of California, Santa Cruz, 2011.
- Adler, Prudence S., Patricia Aufderheide, Brandon Butler y Peter Jaszi. *Code of Best Practices in Fair Use for Academic and Research Libraries*. [S. l.]: Association of Research Libraries, 2012. <http://www.arl.org/storage/documents/publications/code-of-best-practices-fair-use.pdf>.
- Anderson, Jane y Kimberly Christen. “‘Chuck a Copyright on It’: Dilemmas of Digital Return and the Possibilities for Traditional Knowledge Licenses and Labels”. *Museum Anthropology Review* 7, núms. 1-2 (2013): 105-126.
- Arms, Caroline y Carl Fleischhauer. “Digital Formats: Factors for Sustainability, Functionality, and Quality”. Ponencia, Society for Imaging Science and Technology Archiving Conference, 2005. http://memory.loc.gov/ammem/techdocs/di-gform/Formats_IST05_paper.pdf.
- Bagnall, Kate. “Invisible Australians”. *Asian Currents* (abril/mayo de 2012): 14-15.
- Bailey, Jefferson. “I Review 6 Digital Preservation Models So You Don’t Have To”. *Jefferson Bailey* (blog), 26 de abril de 2014. <http://www.jeffersonbailey.com/i-review-6-digital-preservation-models-so-you-dont-have-to/>.
- Bearman, David. *Archival Methods*. Archives and Museum Informatics Technical Report. Vol. 3, núm. 1. Pittsburgh, Pensilvania: Archives & Museum Informatics, 1989.



- Bell, Kim. "Ferguson Protester Who Threw Back Tear Gas Cannister in Iconic Photo Is Charged". *St. Louis Post-Dispatch*, 26 de agosto de 2015. http://www.stltoday.com/news/local/crime-and-courts/ferguson-protester-who-threw-back-tear-gas-cannister-in-iconic/article_437076f9-a6a2-5f03-a2fb-d612611a504f.html.
- Bettivia, Rhiannon. "Encoding Power: The Scripting of Archival Structures in Digital Spaces using the Open Archival Information System (OAIS) Reference Model". Tesis de doctorado, University of Illinois at Urbana-Champaign, 2016.
- Bird, Graeme D. *Multitextuality in the Homeric Iliad. The Witness of the Ptolemaic Papyri*. Hellenic Studies 43. Washington DC: Cambridge, Massachusetts: Center for Hellenic Studies / Harvard University Press, 2010.
- Bogost, Ian. *Alien Phenomenology, or, What It's like to Be a Thing*. Posthumanities 20. Mineápolis: University of Minnesota Press, 2012.
- Brunton, David. "Using Data from Historic Newspapers". *The Signal* (blog), 5 de septiembre de 2017. <https://blogs.loc.gov/thesignal/2017/09/using-data-from-historic-newspapers/>.
- Brylawski, Sam, Maya Lerman, Robin Pike y Kathlin Smith, editores. "ARSC Guide to Audio Preservation". *CLIR Publication*. Washington DC: Association for Recorded Sound Collections / Council on Library and Information Resources / The Library of Congress, 2015. <http://cmsimpact.org/wp-content/uploads/2016/08/ARSC-Audio-Preservation.pdf>.
- Carroll, Laura, Erika Farr, Peter Hornsby y Ben Ranker. "A Comprehensive Approach to Born-Digital Archives". *Archivaria* 72 (2011): 61-92.
- Caswell, Michelle. "Assessing the Use of Community Archives US IMLS RE-31-16-0117-16". Institute of Museum and Library Services. Acceso el 28 de abril de 2021. <https://www.imls.gov/grants/awarded/re-31-16-0117-16>.
- Chan, Sebastian y Aaron Cope. "Collecting the Present: Digital Code and Collections". *MW2014: Museums and the Web 2014*, 2-5 de abril de 2014. <http://mw2014.museumsandtheweb.com/paper/collecting-the-present-digital-code-and-collections/>.
- Christen, Kimberly. "Working Together: Archival Challenges and Digital Solutions in Aboriginal Australia". *SAA Archaeological Recorder* 8, núm. 2 (2008): 21-24.
- Chun, Wendy Hui Kyong. *Control and Freedom. Power and Paranoia in the Age of Fiber Optics*. Cambridge, Massachusetts: MIT Press, 2005.
- "The Conservator-Restorer: A Definition of the Profession". International Council of Museums, Committee for Conservation. Acceso el 28 de abril de 2021. <http://www.icom-cc.org/47/history-of-icom-cc/definition-of-profession-1984>.
- Contaxis, Nicole. "Grateful Med: Personal Computing and User-Friendly Design". *Circulating Now* (blog), 28 de abril de 2016. <https://circulatingnow.nlm.nih.gov/2016/04/28/grateful-med-personal-computing-and-user-friendly-design/>.
- Cook, Terry. "'The Imperative of Challenging Absolutes', in Graduate Archival Education Programs: Issues for Educators and the Profession". *American Archivist* 63, núm. 2 (2000): 380-391.



- Cordell, Ryan y Abby Mullen. “Fugitive Verses”: The Circulation of Poems in Nineteenth-Century American Newspapers”. *American Periodicals* 27, núm. 1 (2017). <http://viraltexts.org/2016/04/08/fugitive-verses/>.
- Dappert, Angela y Adam Farquhar. “Significance Is in the Eye of the Stakeholder”. En *Research and Advanced Technology for Digital Libraries*. Edición de Maristella Agosti, José Borbinha, Sarantos Kapidakis, Christos Papatheodorou y Giannis Tsakonas, 297-308. Berlín; Nueva York: Heidelberg / Springer, 2009. http://link.springer.com/chapter/10.1007/978-3-642-04346-8_29.
- “Deep Blue Preservation and Format Support”. University of Michigan Libraries. Acceso el 28 de abril de 2021. <https://deepblue.lib.umich.edu/static/about/deepbluepreservation.html>.
- Dietrich, Dianne, Julia Kim, Morgan McKeenan y Alison Rhonemus. “How to Party Like It’s 1999: Emulation for Everyone”. *The Code4Lib Journal*, núm. 32 (2016). <http://journal.code4lib.org/articles/11386>.
- Dooley, Jackie. *The Archival Advantage. Integrating Archival Expertise into Management of Born-Digital Library Materials*. Dublín, Ohio: OCLC Research, 2015. <http://www.oclc.org/content/dam/research/publications/2015/oclcresearch-archival-advantage-2015.pdf>.
- Drake, Jarrett M. “Expanding #ArchivesForBlackLives to Traditional Archival Repositories”. *On Archivy*, 27 de junio de 2016. <https://medium.com/on-archivy/expanding-archivesforblacklives-to-traditional-archival-repositories-b88641e2daf6#.6w6jkmgul>.
- Eco, Umberto. *Kant and the Platypus. Essays on Language and Cognition*. Traducción de Alastair McEwen. Nueva York: Harcourt Brace, 2000.
- Emerson, Lori. “Archives, Materiality and the ‘Agency of the Machine’: An Interview with Wolfgang Ernst”. *The Signal* (blog), 8 de febrero de 2013. blogs.loc.gov/the-signal/2013/02/archives-materiality-and-agency-of-the-machine-an-interview-with-wolfgang-ernst/.
- _____. *Reading Writing Interfaces: From the Digital to the Bookbound*. Minneapolis: University of Minnesota Press, 2014.
- Eppink, Jason. “A Brief History of the GIF (so Far)”. *Journal of Visual Culture* 13, núm. 3 (2014): 298-306. doi: 10.1177/1470412914553365.
- Espenschied, Dragan, Isgandar Valizada, Oleg Stobbe, Thomas Liebetaut y Klaus Rechert. “(Re-)Publication of Preserved, Interactive Content—Theresa Duncan CD-ROMs: Visionary Videogames for Girls”. En *Proceedings of the 12th International Conference on Digital Preservation*, 233-234. Chapel Hill: School of Information and Library Science, University of North Carolina at Chapel Hill, 2015. <http://phaidra.univie.ac.at/o:429524>.
- Fino-Radin, Ben. “Rhizome Artbase: Preserving Born Digital Works of Art”. Conferencia, Digital Preservation 2012, 24-26 de julio de 2012. http://digitalpreservation.gov/meetings/documents/ndiipp12/DigitalCulture_fino-radin_DP12.pdf.
- _____. “Take a Picture, It’ll Last Longer...”. *Ben Fino-Radin’s Blog*, 28 de agosto de 2012. <http://notepad.benfinoradin.info/2012/08/28/take-a-picture/>.



- Fino-Radin, Ben. "The Web Browser As Aesthetic Framework: Why Digital Art Today Looks Different". *Vice*, 8 de mayo de 2012. https://creators.vice.com/en_us/article/digart-the-web-browser-as-aesthetic-framework-why-digital-art-today-looks-different.
- Fraistat, Neil. *Shared Horizons: Data, BioMedicine, and the Digital Humanities*. College Park: Digital Repository, University of Maryland, 2013. <http://drum.lib.umd.edu/handle/1903/14721>.
- Gitelman, Lisa. *Always Already New. Media, History, and the Data of Culture*. Cambridge, Massachusetts: MIT Press, 2006.
- _____. *Paper Knowledge. Toward a Media History of Documents*. Durham; Londres: Duke University Press Books, 2014.
- Greene, Mark y Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing". *The American Archivist* 68, núm. 2 (2005): 208-263.
- Grinspoon, David. *Earth in Human Hands. Shaping Our Planet's Future*. Nueva York: Grand Central, 2016.
- Hagan, William. "Archival Captive—The American Indian". *American Archivist* 41, núm. 2 (1978): 135-142.
- Hedstrom, Margaret, Christopher Lee, Judith Olson y Clifford Lampe. "The Old Version Flickers More": Digital Preservation from the User's Perspective". *American Archivist* 69, núm. 1 (2006): 159-187.
- Houghton, Bernadette. "Trustworthiness: Self-Assessment of an Institutional Repository against ISO 16363-2012". *D-Lib Magazine* 21, núms. 3-4 (2015). DOI: 10.1045/march2015-houghton.
- Hunn, David. "Subject of Iconic Photo Speaks of Anger, Excitement". *St. Louis Post-Dispatch*, 24 de agosto de 2014. http://www.stltoday.com/news/local/crime-and-courts/subject-of-iconic-photo-speaks-of-anger-excitement/article_3076e398-2c7b-5706-9856-784c997d0a52.html.
- "Hyperspectral Imaging by Library of Congress Reveals Change Made by Thomas Jefferson in Original Declaration of Independence Draft" [2 de julio de 2010]. Office of Communications, and Library of Congress Office of Communications. Acceso el 28 de abril de 2021. <https://www.loc.gov/item/prn-10-161/analysis-reveals-changes-in-declaration-of-independence/2010-07-02/>.
- Jules, Bergis. "Documenting the Now: #Ferguson in the Archives". *OnArchivy*, 8 de abril de 2015. <https://medium.com/on-archivy/documenting-the-now-ferguson-in-the-archives-adcdbe1d5788>.
- Kittler, Friedrich A. *Gramophone, Film, Typewriter*. Traducción de Michael Wutz y Geoffrey Winthrop-Young. Stanford, California: Stanford University Press, 1999.
- Kirschenbaum, Matthew. *Mechanisms. New Media and the Forensic Imagination*. Cambridge, Massachusetts: MIT Press, 2008.
- _____. "Software, It's a Thing". *Medium*, 25 de julio de 2014. <https://medium.com/@mkirschenbaum/software-its-a-thing-a550448d0ed3>.



- Jackson, Steven J. "Rethinking Repair". En *Media Technologies Essays on Communication, Materiality, and Society*. Edición de Tarleton Gillespie, Pablo J. Boczkowski y Kirsten A. Foot. Cambridge, Massachusetts: MIT Press, 2014.
- Krajewski, Markus. *Paper Machines. About Cards & Catalogs, 1548-1929*. History and Foundations of Information Science. Cambridge, Massachusetts: MIT Press, 2011.
- Kriesberg, Adam M. "The Changing Landscape of Digital Access: Public-Private Partnerships in US State and Territorial Archives". Tesis de doctorado, University of Michigan, 2015.
- Kruse, Warren G. y Jay G. Heiser. *Computer Forensics. Incident Response Essentials*. Boston: Addison-Wesley Professional Education, 2002.
- Lavoie, Brian F. *The Open Archival Information System (OAIS) Reference Model: Introductory Guide. DPC Technology Watch Report 14-02 October 2014*. Gran Bretaña: Digital Preservation Coalition, 2014. <http://www.dpconline.org/docman/technology-watch-reports/1359-dpctw14-02/file>.
- Lee, Christopher. "Digital Curation as Communication Mediation". En *Handbook of Technical Communication*. Edición de Alexander Mehler y Laurent Romary, 507-530. Berlín; Boston, Massachusetts: De Gruyter Mouton, 2012.
- LeFurgy, Bill. "Life Cycle Models for Digital Stewardship". *The Signal* (blog), 21 de febrero de 2012. <http://blogs.loc.gov/thesignal/2012/02/life-cycle-models-for-digital-stewardship/>.
- Lorang, Elizabeth M., Leen-Kiat Soh, Maanas Varma Datla y Spencer Kulwicki. "Developing an Image-Based Classifier for Detecting Poetic Content in Historic Newspaper Collections". *D-Lib* 21, núms. 7-8 (2015). <http://digitalcommons.unl.edu/librarianscience/340/>.
- Lowood, Henry. "Memento Mundi: Are Virtual Worlds History?". *iPRES 2009. The Sixth International Conference on the Preservation of Digital Objects. Proceedings*. 121-128. S. l.: California Digital Library, 2009. <http://escholarship.org/uc/item/2gs3p6jx>.
- Manovich, Lev. "Database as a Genre of New Media". *AI & Society* 14, núm. 2 (2000): 176-183. doi: <https://doi.org/10.1007/BF01205448>.
- _____. *The Language of New Media*. Cambridge, Massachusetts: MIT Press, 2002.
- _____. *Software Takes Command. Extending the Language of New Media*. International Texts in Critical Media Aesthetics 5. Nueva York: Bloomsbury, 2013.
- Marshall, Catherine C. "Digital Copies and a Distributed Notion of Reference in Personal Archives". En *Digital Media: Technological and Social Challenges of the Interactive World*. Edición de Megan Alicia Winget y William Aspray, 89-115. Lanham: Scarecrow Press, 2011.
- McDonough, Jerome, Robert Olendorf, Matthew Kirschenbaum, Kari Kraus, Doug Reside, Rachel Donahue, Andrew Phelps, Christopher Egert, Henry Lowood y Susan Rojo. *Preserving Virtual Worlds. Final Report*. [S. l.]: Library of Congress, 2010. <http://hdl.handle.net/2142/17097>.
- McNeill, Lynne S. *Folklore Rules. A Fun, Quick, and Useful Introduction to the Field of Academic Folklore Studies*. Logan: University Press of Colorado, 2013.



- Milligan, Ian, Nick Ruest y Jimmy Lin. "Content Selection and Curation for Web Archiving: The Gatekeepers vs. the Masses". *Digital Libraries (JCDL), 2016 IEEE/ACM Conference on Digital Libraries* (2016), 107-110. <http://ieeexplore.ieee.org/abstract/document/7559571/>.
- "Minimal Computing". Global Outlook: Digital Humanities. Acceso el 28 de abril de 2021. www.globaloutlookdh.org/minimal-computing/.
- Mir, Rebecca y Trevor Owens. "Modeling Indigenous Peoples: Unpacking Ideology in Sid Meier's Colonization". En *Playing with the Past: Digital Games and the Simulation of History*. Edición de Matthew Wilhelm Kapel y Andrew B. R. Elliott, 91-106. Nueva York; Londres: Bloomsbury, 2013. https://books.google.com/books?hl=en&lr=&id=gZPFAgAAQBAJ&oi=fnd&pg=PA91&dq=mir+owens+indigenous&ots=B7s8anU04z&sig=_tGtLeWJPCPxunsXoVFUycIJLrM.
- Molloy, Mark. "Palestinians Tweet Tear Gas Advice to Protesters in Ferguson". *The Telegraph*, 15 de agosto de 2014. <http://www.telegraph.co.uk/news/worldnews/northamerica/usa/11036190/Palestinians-tweet-tear-gas-advice-to-protesters-in-Ferguson.html>.
- Montfort, Nick. "Continuous Paper: MLA". Conferencia, Modern Language Association, 28 de diciembre de 2004. http://nickm.com/writing/essays/continuous_paper_mla.html.
- _____. e Ian Bogost. *Racing the Beam. The Atari Video Computer System*. Platform Studies. Cambridge, Massachusetts: MIT Press, 2009.
- Morozov, Evgeny. *To Save Everything, Click Here. The Folly of Technological Solutionism*. Nueva York: Public Affairs, 2013.
- Nakamura, Lisa. *Digitizing Race. Visual Cultures of the Internet*. Electronic Mediations 23. Mineápolis: University of Minnesota Press, 2008.
- Nowviskie, Bethany. "On Capacity and Care". Discurso, "National Endowment for the Humanities-Digital Humanities Project Directors' Meeting", septiembre de 2015.
- Ongondo, F. O., Ian D. Williams y Tom Cherrett. "How Are WEEE Doing? A Global Review of the Management of Electrical and Electronic Wastes". *Waste Management* 31, núm. 4 (2011): 714-730.
- Owens, Trevor. "Creepypastas, Memes, Lolspeak & Boards: The Scope of a Digital Culture Web Archive". *Folklife Today* (blog), 26 de septiembre de 2014. blogs.loc.gov/folklife/2014/09/scoping-a-digital-culture-web-archive/.
- _____. "Exhibiting .Gifs: An Interview with Curator Jason Eppink". *The Signal* (blog), 2 de junio de 2014. <http://blogs.loc.gov/digitalpreservation/2014/06/exhibiting-gifs-an-interview-with-curator-jason-eppink/>.
- _____. "Life-Saving: The National Software Reference Library". *The Signal* (blog), 4 de mayo de 2012. <http://blogs.loc.gov/thesignal/2012/05/life-saving-the-national-software-reference-library/>.
- _____. "Mecha-Archivists: Envisioning the Role of Software in the Future of Archives". *Trevor Owens: User Centered Digital History*, 27 de mayo de 2014. <http://www.trevorowens.org/2014/05/mecha-archivists-envisioning-the-role-of-software-in-the-future-of-archives/>.



- Owens, Trevor. "Open Source Software and Digital Preservation: An Interview with Bram van Der Werf of the Open Planets Foundation". *The Signal* (blog), 4 de abril de 2012. <http://blogs.loc.gov/thesignal/2012/04/open-source-software-and-digital-preservation-an-interview-with-bram-van-der-werf-of-the-open-planets-foundation/>.
- _____. "Personal Stories, Storage Media, and Veterans History: An Interview with Andrew Cassidy-Amstutz". *The Signal* (blog), 6 de marzo de 2014. <http://blogs.loc.gov/thesignal/2014/03/personal-stories-storage-media-and-veterans-history-an-interview-with-andrew-cassidy-amstutz/>.
- _____. "Pixelated Commemorations: 4 In Game Monuments and Memorials". *Play the Past*, 18 de junio de 2014. <http://www.playthepast.org/?p=4811>.
- _____. "Sagan's Thinking and Writing Process". En *Finding Our Place in the Cosmos: From Galileo to Sagan and Beyond*. Washington DC: Library of Congress, 2014. <https://www.loc.gov/collections/finding-our-place-in-the-cosmos-with-carl-sagan/articles-and-essays/carl-sagan-and-the-tradition-of-science/sagans-thinking-and-writing-process/>.
- _____. "What Do You Mean by Archive? Genres of Usage for Digital Preservers". *The Signal* (blog), 27 de febrero de 2014. <http://blogs.loc.gov/thesignal/2014/02/what-do-you-mean-by-archive-genres-of-usage-for-digital-preservers/>.
- Phillips, Megan, Jefferson Bailey, Andrea Goethals y Trevor Owens. "The NDSA Levels of Digital Preservation: An Explanation and Uses". Presentación, IS&T Archiving Conference, 2-5 de abril de 2013. http://www.digitalpreservation.gov/documents/NDSA_Levels_Archiving_2013.pdf.
- "Popcorn GIFs". Know Your Meme. Acceso el 28 de abril de 2021. <http://knowyourmeme.com/memes/popcorn-gifs>.
- Reich, Victoria y David Rosenthal. "Distributed Digital Preservation: Lots of Copies Keep Stuff Safe". En *Proceedings Indo-US Workshop on International Trends in Digital Preservation*. 51-55. Pune: Human Centered Design & Computing C-CAD, 2009. <https://web.archive.org/web/20160507064543/http://lockss.org/locksswiki/files/ReichIndiaFinal.pdf>.
- Reside, Doug. "'No Day But Today': A Look at Jonathan Larson's Word Files". *New York Public Library Blog*, 22 de abril de 2011. <http://www.nypl.org/blog/2011/04/22/no-day-today-look-jonathan-larsons-word-files>.
- Richard Vijgen. *Deleted Cities* [2011]. Rhizome. Acceso el 28 de abril de 2021. <http://rhizome.org/art/artbase/artwork/deleted-cities/>.
- Rinehart, Richard y Jon Ippolito, editores. *Re-Collection. Art, New Media, and Social Memory*. Cambridge, Massachusetts: MIT Press, 2014.
- Rose, Matthew y Jonathan Star. *Using Scenarios to Explore Climate Change: A Handbook for Practitioners*. [Washington DC]: National Park Service, U. S. Department of the Interior, 2013.
- Rosenthal, David. "Storage Will Be a Lot Less Free than It Used to Be". *DSHR's Blog*, 1 de octubre de 2012. <https://blog.dshr.org/2012/10/storage-will-be-lot-less-free-than-it.html>.



- Rosenthal, David, Kris Carpenter y Krishna Kant. "Green Bytes: Sustainable Approaches to Digital Stewardship". Panel de discusión, "Library of Congress Digital Preservation 2013 Meeting", 23-25 de julio de 2013.
- Rosenthal, David, Daniel C. Rosenthal, Ethan L. Miller, Ian F. Adams, Mark W. Storer y Erez Zadok. "The Economics of Long-Term Digital Storage". Ponencia, "The Memory of the World in the Digital Age: Digitization and Preservation UNESCO Conference", 26-28 de septiembre de 2012.
- Rudersdorf, Amy. "Proof Is in the Standards Pudding". *AVP Blog*, 21 de junio de 2016. <https://www.weareavp.com/the-proof-is-in-the-standards-pudding/>.
- Russell, Andrew y Lee Vinsel. "Hail the Maintainers". *Aeon*, 7 de abril de 2017. <https://aeon.co/essays/innovation-is-overvalued-maintenance-often-matters-more>.
- Salo, Dorothea. "Innkeeper at the Roach Motel". *Library Trends* 57, núm. 2 (2008): 98-123.
- Saylor, Nicole. "Computing Culture in the AFC Archive". *Folklife Today* (blog), 8 de enero de 2014. <https://blogs.loc.gov/folklife/2014/01/computing-culture-in-the-afc-archive/>.
- Scott, Jason. "Archiveteam! The Geocities Torrent". *ASCII by Jason Scott* (blog), 27 de octubre de 2010. <http://ascii.textfiles.com/archives/2720>.
- Section 108 Study Group. *The Section 108 Study Group Report*. Washington DC: The Library of Congress, 2008. <http://www.section108.gov/docs/Sec108StudyGroupReport.pdf>.
- Sharpless, Rebecca. "The History of Oral History". En *History of Oral History. Foundations and Methodology*. Edición de Thomas L. Charlton, Lois E. Myers y Rebecca Sharpless, 9-32. Lanham: AltaMira Press, 2007.
- Sheldon, Madeline. "Analysis of Current Digital Preservation Policies: Archives, Libraries, and Museums". *The Signal* (blog), 13 de agosto de 2013. <https://blogs.loc.gov/thesignal/2013/08/analysis-of-current-digital-preservation-policies-archives-libraries-and-museums/>.
- Sherratt, Tim. "The Real Face of White Australia". *Discontents* (blog), 21 de septiembre de 2011. <http://discontents.com.au/the-real-face-of-white-australia/>.
- Shulgin, Alexei. *Form Art* [1997]. Rhizome. Acceso el 28 abril de 2021. <http://rhizome.org/art/artbase/artwork/form-art/>.
- Simon, Nina. *The Art of Relevance*. Santa Cruz, California: Museum 2.0, 2016.
- Skinner, Katherine y Martin Halbert. "The MetaArchive Cooperative: A Collaborative Approach to Distributed Digital Preservation". *Library Trends* 57, núm. 3 (2009): 371-392.
- Smigel, Libby, Martha Goldstein y Elizabeth Aldrich. *Documenting Dance. A Practical Guide*. Washington DC: Dance Heritage Coalition, 2006. <https://web.archive.org/web/20180417151147/http://www.danceheritage.org/DocumentingDance.pdf>.
- Stefan, Paula De, Carl Fleischhauer, Andrea Goethals, Michael Kjörling, Nick Krabbenhoef, Chris Lacinak, Jane Mandelbaum. "Checking Your Digital Content: What Is Fixity, and When Should I Be Checking It?". Reporte de trabajo, National Digital Stewardship Alliance, 2014. <http://hdl.loc.gov/loc.gdc/lcpub.2013655117.1>.



- Sterne, Jonathan. *MP3. The Meaning of a Format. Sign, Storage, Transmission*. Durham: Duke University Press, 2012.
- Strauss, Benjamin. "What Does U. S. Look Like with 10 Feet of Sea Level Rise?". *Climate Central*, 13 de mayo de 2014. <http://www.climatecentral.org/news/us-with-10-feet-of-sea-level-rise-17428>.
- Swan, Deanne W., Justin Grimes y Timothy Owens. *The State of Small and Rural Libraries in the United States*. Research Brief 5. Washington DC: Institute of Museum and Library Services, 2013. <https://www.ims.gov/sites/default/files/publications/documents/brief201305.pdf>.
- Tansey, Eira. "Archival Adaptation to Climate Change". *Sustainability: Science, Practice & Policy* 11, núm. 2 (2015): 45-56.
- Thacker, Eugene. "Protocol Is as Protocol Does". En *Protocol. How Control Exists after Decentralization*. Edición de Alexander R. Galloway, xi-xxii. Cambridge, Massachusetts: MIT Press, 2006.
- "The 2015 Pulitzer Prize-Winning Photographs from the Post-Dispatch". *St. Louis Post-Dispatch*, 20 de abril de 2015. http://www.stltoday.com/news/local/metro/the-pulitzer-prize-winning-photographs-from-the-post-dispatch/collection_7a5793c3-9a55-534c-b84a-a27d6d08ef5f.html.
- "Thesaurus Linguae Graecae—History". Thesaurus Linguae Graecae Project. Acceso el 28 de abril de 2021. <https://web.archive.org/web/20120129001615/http://www.tlg.uci.edu/about/history.php>.
- Tillett, Barbara. "A Conceptual Model for the Bibliographic Universe". *Technicalities* 25, núm. 5 (2003). <http://www.loc.gov/cds/downloads/FRBR.PDF>.
- Tomasello, Michael. *The Cultural Origins of Human Cognition*. Boston, Massachusetts: Harvard University Press, 2009.
- Tyrrell, Ian R. *Historians in Public. The Practice of American History, 1890-1970*. Chicago: University of Chicago Press, 2005.
- Vaidhyanathan, Siva. *The Googlization of Everything*. Berkeley: University of California Press, 2012.
- Walter, Chip. "Kryder's Law". *Scientific American* (2005): 32-33.
- Web, Colin, David Pearson y Paul Koerben. "'Oh, You Wanted Us to Preserve That?!' Statements of Preservation Intent for the National Library of Australia's Digital Collections". *D-Lib Magazine* 19, núms. 1-2 (2013). DOI: 10.1045/january2013-webb.
- Werner, Sarah. "Where Material Book Culture Meets Digital Humanities". *Journal of Digital Humanities* 1, num. 3 (2012). <http://journalofdigitalhumanities.org/1-3/where-material-book-culture-meets-digital-humanities-by-sarah-werner/>.
- Whalen, Zach. "Close Enough". Presentación, Modern Language Association, 8 de enero de 2012. <http://www.zachwhalen.net/posts/mla-2012-close-enough/>.
- Yeo, Geoffrey. "'Nothing Is the Same as Something Else': Significant Properties and Notions of Identity and Originality". *Archival Science* 10, núm. 2 (2010): 85-116.
- Zinn, Howard. "Secrecy, Archives, and the Public Interest". *Midwestern Archivist* 2, núm. 2 (1977): 14-26.

La teoría y el oficio de la preservación digital,

editado por el Instituto de Investigaciones Bibliográficas de la UNAM. Versión PDF se terminó en agosto de 2023.

Para su composición y formación tipográfica se utilizó la fuente Baskerville en 9, 10, 11, 14 y 20 puntos.

Departamento Editorial del IIB

Coordinación editorial Josué Humberto Brocca Tovar Kuri

Corrección de estilo María Fernanda Baroco Gálvez

Cuidado editorial María José Ramírez Herrera

Asistencia editorial José Leonardo Hernández López

Diseño editorial y formación Hilda Maldonado

Este libro ofrece una introducción básica a los problemas y prácticas de la preservación digital y a la naturaleza de la información y los medios digitales, con el objetivo de posibilitar avances significativos y prácticos encaminados a mitigar los riesgos más apremiantes de la pérdida digital, a la vez que sirve como un marco teórico para abordar la preservación digital deliberada e intencionalmente en cuanto a un campo con múltiples vertientes.

Si bien tiene una inclinación práctica, el ejemplar que el lector tiene en sus manos no es un manual, tampoco es un conjunto de instrucciones paso a paso, más bien está pensado como un punto de referencia para desarrollar y perfeccionar el oficio de un profesional de la preservación digital, pero también para cualquier persona interesada en comenzar a practicarla en cualquier ámbito.

Esta primera edición al español de la obra de Trevor Owens representa un esfuerzo por franquear las barreras idiomáticas que entorpecen el acceso equitativo al conocimiento en un ámbito académico donde predomina la comunicación en inglés. La traducción de este libro se realizó como parte de una experiencia colaborativa a través de herramientas y plataformas digitales, con el propósito de poner a disposición de la comunidad hispanohablante una aportación significativa para el campo de la preservación digital, así como con la intención de auxiliar a las personas que se preocupan y ocupan por esta importante tarea.



INSTITUTO DE
INVESTIGACIONES
BIBLIOGRÁFICAS

