

TEMA 7

El Modelo de Ciclo de Vida del Archivado Web

por Kristine Hanna

Jefe del Servicio de Archivado del Internet Archive<https://archive.org/about/bios.php>

INTRODUCCIÓN

Desde hace más de una década, las herramientas tecnológicas de archivado web han registrado una evolución constante. Sin embargo, aún no han surgido buenas prácticas ni un modelo común de archivado web. Este Modelo de Ciclo de Vida del Archivado Web constituye un intento de incorporar las facetas tecnológicas y programáticas del archivado web a un marco que resulte relevante para cualquier organización que desee archivar la web. Archive-It, el servicio líder en la comunidad de archivo web, desarrolló el modelo a partir de su trabajo con instituciones encargadas de conservar el patrimonio en todo el mundo.

El Internet Archive (Archivo de Internet) lleva archivando la web desde 1996. En 2002, el Internet Archive lanzó Heritrix, el rastreador web de código abierto. En 2009, el archivo resultante del rastreador Heritrix, el archivo WARC, fue adoptado como norma ISO para archivado web, demostrando así la preponderancia de los programas de archivado activo de la web y la importancia del propio rastreador. A principios de 2006, el Internet Archive puso en marcha el servicio de archivo Archive-It (www.archive-it.org) con trece instituciones piloto asociadas. Archive-It consiste en un servicio de suscripción para el archivado web que ayuda a las instituciones asociadas a recopilar, crear y gestionar colecciones de origen digital. La base de socios ha ido creciendo pro-

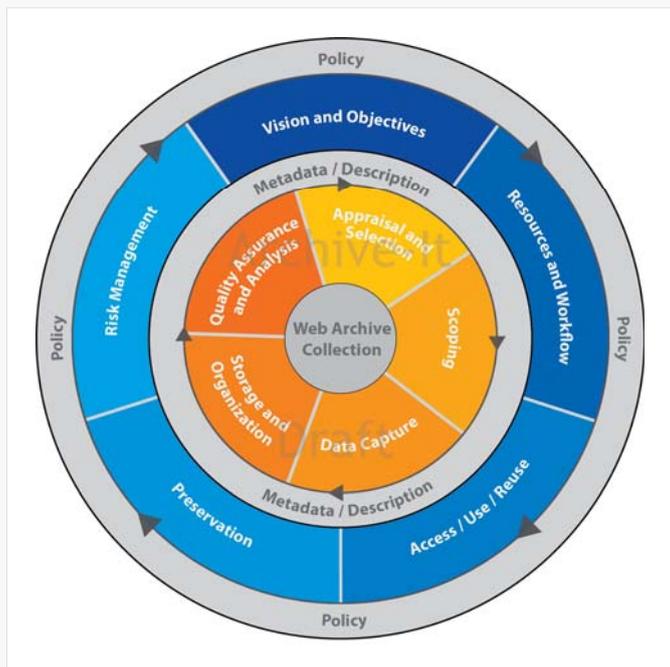
gresivamente desde su creación y, en enero de 2013, contaba con 237 socios en 46 estados norteamericanos y 15 países.

A pesar del aumento del número de programas de archivado web, son muchas las instituciones que siguen teniendo dificultades para desarrollar buenas prácticas y metodologías a la hora de cumplir sus objetivos. Esta dificultad se debe en parte a la evolución constante de la tecnología web, ya que puede resultar complicado archivar determinados tipos de contenido de manera eficaz. La adopción de políticas contradictorias y sujetas a cambios por parte de diversas partes interesadas y los cambios en las estructuras organizativas y respon-

sabilidades laborales constituyen un obstáculo adicional a la hora de establecer buenas prácticas. Asimismo, algunas organizaciones interesadas no han asumido plenamente que el archivado de la web es esencial para sus actividades de conservación digital; por tanto, la financiación sigue siendo limitada o inexistente.

Con el objeto de resolver la ausencia de buenas prácticas normalizadas y de reforzar la sensibilización sobre la importancia del archivado de la web como actividad fundamental para la conservación

Archive-It es un servicio de archivo web que sigue el Modelo de Ciclo de Vida del Archivado Web



GESTIÓN DEL CICLO DE VIDA DEL ARCHIVADO WEB

GLOSARIO

Policy	Política
Vision and Objectives	Visión y objetivos
Resources and Workflow	Recursos y flujo de trabajo
Access/Use/Reuse	Acceso, utilización y reutilización
Preservation	Conservación
Risk Management	Gestión del riesgo
Metadata/Description	Metadatos/descripción
Quality Assurance and Analysis	Aseguramiento y análisis de la calidad
Appraisal and Selection	Valoración y selección
Scoping	Definición del alcance
Data Capture	Captura de datos
Storage and Organization	Almacenamiento y organización
Web Archive Collection	Colección de archivos web

digital, el equipo de Archive-It ha desarrollado el Modelo de Ciclo de Vida del Archivado Web (WALCM, por sus siglas en inglés). Este modelo se basa en las experiencias de dicho equipo y lo aprendido por innumerables instituciones asociadas, incluido el estudio exhaustivo de casos concretos de seis de estas instituciones. El WALCM constituye un intento de representar flujos de trabajo comunes y crear un modelo cuantificable que sirva de referencia a las organizaciones para desarrollar o mejorar sus programas de archivado de la web.

DESARROLLAR EL MODELO DE CICLO DE VIDA DEL ARCHIVADO WEB

El equipo de Archive-It desarrolló el modelo de forma orgánica, aprovechando la información y el aprendizaje adquiridos por organizaciones que archivan la web. La mayoría de estas organizaciones utilizan el servicio de Archive-It para archivar contenidos web para sus organizaciones. Estas instituciones asociadas facilitan al Internet Archive información basada en el uso que hacen del servicio y la comunican mediante correo electrónico, teléfono y conversaciones presenciales en conferencias y reuniones de socios. Además, se obtiene más información de carácter formal gracias a las presentaciones de los socios en las conferencias y las encuestas que diseña el personal de Archive-It, así como la documentación y las presentaciones de los socios en las conferencias sobre cómo se enfrentan ellos y sus colegas a los retos que plantea el archivado web.

El equipo de Archive-It esbozó la primera iteración del Modelo de Ciclo de Vida del Archivado Web. Este diseño preliminar se distribuyó a un subgrupo de socios de Archive-It que aportó información sobre elementos omitidos o superfluos y sobre la presentación visual del modelo. A continuación, el equipo de Archive-It incorporó todos estos datos a un modelo gráfico más atractivo y lo envió a todos sus socios para obtener una opinión general. Con esta información se desarrolló otro nuevo diseño y la versión resultante del modelo que se expone en este documento. La información que se recoge aquí procede también de intensos intercambios de correo electrónico y entrevistas telefónicas con seis socios de Archive-It que tuvieron lugar entre abril y julio de 2012. Estas instituciones son la Universidad de Columbia, la Universidad de Alberta, la Biblioteca Estatal de Montana, la Biblioteca Estatal de Carolina del Norte, los Archivos Estatales de Carolina del Norte y la Universidad de Creighton. La información que se recoge en este documento procede asimismo de una encuesta a los socios de Archive-It realizada en agosto de 2012.

EXPLICACIÓN DEL MODELO

El modelo constituye un intento de extraer los distintos pasos y fases por los que pasa una institución durante el desarrollo y la gestión de su programa de archivado web. Aunque el modelo se desglosa en pasos individuales, cada acción no es independiente. Archive-It considera que los pasos y las fases están relacionados y muy solapados.

El modelo tiene forma circular para sugerir la naturaleza repetitiva de los pasos en el ciclo de vida. Al avanzar por cada paso, los usuarios terminan volviendo al punto de partida o repiten determinados pasos, dependiendo de cuáles sean sus tareas. Por ejemplo, el proceso puede volver a iniciarse cuando una institución incorpora nuevos sitios web a una colección ya existente o crea una colección totalmente nueva. El modelo se compone de círculos concéntricos, lo que sugiere que estos ciclos repetitivos pertenecen a un proceso de mayor envergadura.

En el nivel más externo del ciclo de vida se encuentra la banda correspondiente a la política.

Prácticamente todos los aspectos del archivado web conllevan algún tipo de decisión política. Estas decisiones políticas pueden implicar el desarrollo de una nueva política específica del archivado web o la adaptación de una política ya existente a nuevas situaciones. Al englobar los pasos del ciclo de vida dentro de una banda de política, el modelo representa visualmente la naturaleza siempre presente de la formulación de políticas. En una segunda banda, el modelo representa igualmente los metadatos y la descripción. Archive-It decidió incorporar los metadatos como banda y no como segmento de la rueda para hacer hincapié en que la creación, importación y exportación de metadatos pueden realizarse como parte de otras actividades del ciclo de vida.

El modelo sigue las distintas fases por las que pasa una institución durante el desarrollo de su programa de archivado web

El círculo azul comprendido justo dentro de la banda de política representa las decisiones de alto nivel a las que se enfrenta una institución a la hora de establecer y gestionar su programa de archivado web. A continuación se definen los pasos individuales, que se tratarán con más detalle posteriormente en este documento.

- **Visión y objetivos:** aquí las instituciones esclarecen los objetivos de su programa de archivado web.
- **Recursos y flujo de trabajo:** las instituciones examinan los recursos de que disponen, como los recursos financieros, la especialización, el personal, los posibles colaboradores, entre otros, con el objeto de determinar el modo de proceder a la hora de desarrollar o modificar su programa de archivado web.
- **Acceso, utilización y reutilización:** las instituciones deciden si facilitan el acceso a sus colecciones y la manera de hacerlo, además de monitorizar cómo utilizan sus usuarios el contenido.
- **Conservación:** las instituciones deciden cómo desean conservar los datos que recopilan en sus actividades de archivado web. Se incluyen aquí archivos WARC, metadatos y X.
- **Gestión del riesgo:** cuando las instituciones se plantean cómo enfocar el riesgo a la hora de desarrollar un programa de archivado web, analizan los derechos de autor y los permisos, además del acceso.

El círculo interior de color naranja describe las tareas cotidianas que implica la actividad de archivar la web. Estas tareas comprenden las siguientes actividades:

- **Valoración y selección:** las instituciones deciden específicamente los sitios web que desean recopilar.

- **Definición del alcance:** las instituciones pueden optar por archivar un sitio web de manera parcial o total, incluso dominios web completos.
- **Captura de datos:** aquí las instituciones ajustan el modo en que desean capturar los datos decidiendo la frecuencia de rastreo y los tipos de ficheros que deben o no archivers. Las fases de definición del alcance y de captura de datos del ciclo de vida suelen solaparse porque conllevan actividades y decisiones similares.
- **Almacenamiento y organización:** este paso comprende un plan de almacenamiento provisional o a largo plazo de los datos archivados. Para algunas instituciones, la fase de almacenamiento y organización del ciclo de vida también puede constituir sus actividades de conservación.
- **Aseguramiento y análisis de la calidad:** aquí las instituciones examinan el material archivado y en qué medida la colección resultante satisface los objetivos que se fijaron al comienzo del ciclo de vida.

En el centro del ciclo de vida se encuentra la colección propiamente dicha, el contenido web archivado. Estos datos constituyen el resultado final de todos los pasos anteriores y son el material que se va a conservar. La captura y conservación de colecciones de datos constituyen el núcleo de todas las actividades de archivado web y, por tanto, el centro de este modelo.

EL MODELO DE CICLO DE VIDA DEL ARCHIVADO WEB

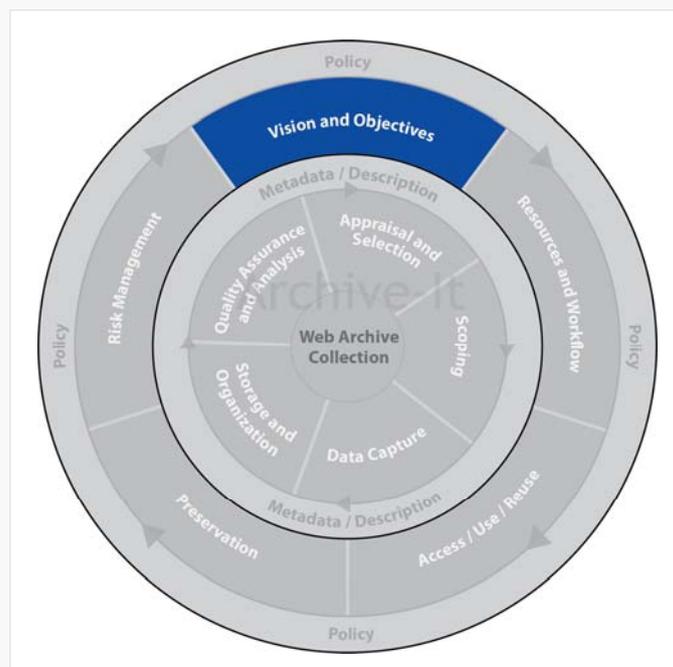
1. El círculo exterior

1a. Visión y objetivos

A la hora de establecer una visión y un objetivo de archivado web, una institución debe preguntarse por qué ha decidido archivar la web, qué desea conse-

guir con ello y qué relación guardan estos pasos con la misión más amplia de la institución. Este paso del ciclo tiene lugar principalmente cuando las instituciones planifican su programa inicialmente; no obstante, las instituciones tienden a reexaminar y redefinir sus objetivos de archivado web a lo largo de toda la duración del programa. Estos periodos de revisión pueden derivarse de un estímulo específico, como un cambio de recursos, o puede tratarse de una cuestión continua que se plantea en relación o junto con otras políticas de recopilación.

Las instituciones de conservación patrimonial deciden archivar la web por motivos muy distintos, dependiendo de sus propios mandatos institucionales y de los objetivos de sus grupos de interés. Hay instituciones que deciden archivar la web porque creen que ese contenido web específico corre el riesgo de desaparecer y, por tanto, debe capturarse y mantenerse accesible, especialmente en caso de acontecimientos espontáneos que cambian rápidamente,



EL CÍRCULO EXTERIOR - VISIÓN Y OBJETIVOS

como catástrofes de origen natural o humano, sublevaciones políticas y actos conmemorativos de personalidades. Otras instituciones tienen el mandato de archivar publicaciones específicas que solo existen en formato digital, como pueden ser programas de cursos universitarios o informes y publicaciones de agencias estatales. Además, hay instituciones que tienen la obligación legal de archivar todos los registros oficiales que produzcan dentro de su propio ámbito, creando así un registro histórico de la presencia web de su institución a lo largo del tiempo. Por otra parte, hay instituciones que consideran el archivado web una ampliación de su política general de desarrollo de colecciones o sus programas de conservación digital, por lo que podrían archivar contenidos web que mejoren o complementen asuntos que ya están recibiendo atención en sus actividades tradicionales de recopilación. Los investigadores y académicos comprenden la importancia que reviste crear un archivo web temático sobre una materia o un asunto específico que comprenda diferentes puntos de vista y

comentarios sociales procedentes de la creciente influencia de las redes sociales, como son los mensajes de Twitter (*tweets*),

los blogs, las entradas (*posts*) y los comentarios en las redes. Algunas instituciones tienen distintos objetivos y, por tanto, establecen diversas colecciones para cumplir cada uno de esos objetivos. Independientemente de la visión específica que tenga cada programa de archivado web, la presente visión configura muchas de las políticas y decisiones que se adoptan en pasos posteriores del ciclo de vida del archivado web.

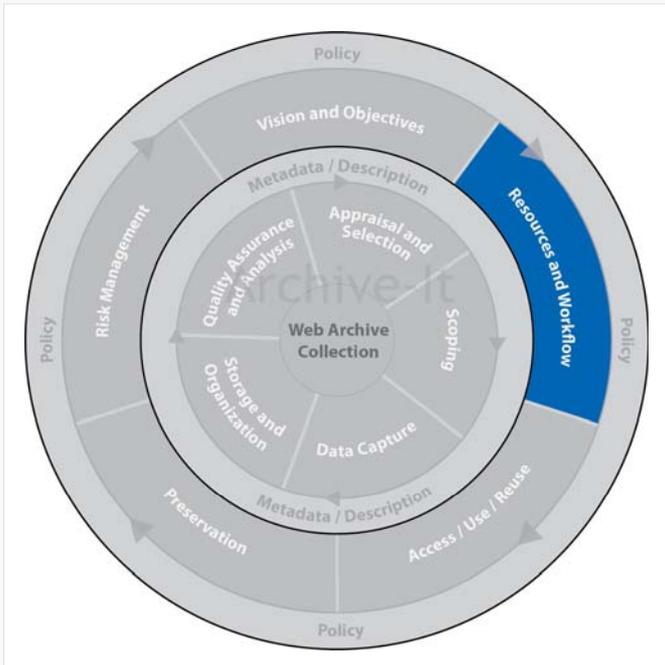
Por ejemplo, la Biblioteca de la Universidad de Columbia trabaja con Archive-It desde 2008 recopilando contenidos web procedentes de diversos ámbitos. En primer lugar, la biblioteca captura el dominio web de la Universidad de Columbia en coordinación con University Archives. En segundo lugar, la biblioteca cuenta con otras colecciones elaboradas

Las instituciones archivan la web por motivos y con objetivos muy distintos siguiendo una estrategia de conservación digital

en torno a temas y asuntos específicos. Estos temas comprenden los derechos humanos en el mundo, la conservación histórica y la planificación urbana y las instituciones religiosas de la ciudad de Nueva York. Estas colecciones de origen digital complementan y suplementan las actividades físicas de recopilación ya existentes de la biblioteca. La Universidad de Columbia describe su objetivo general de archivado web como «crear que el contenido disponible de forma gratuita constituye una fuente de contenidos cada vez más importante y necesaria para la investigación actual y futura que aún no se ha integrado en los modelos de desarrollo de las colecciones de bibliotecas académicas» (correspondencia personal con Alex Thurman y Tessa Fallon, mayo de 2012).

Al igual que la Universidad de Columbia, la Universidad de Alberta también se percató de que no estaba capturando material digital y que necesitaba incorporar el archivado web a su estrategia de conservación digital. Sin embargo, la Universidad no tenía una visión tan clara inicialmente. Al principio, la Universidad de Alberta heredó más de 80 sitios web de una organización sin ánimo de lucro que se había quedado sin financiación. Al darse cuenta de que el alojamiento de estos sitios web requeriría importantes recursos, la Universidad adoptó un enfoque de «archivado» que, en su opinión, sería un modo más sostenible de custodiar el contenido. De este modo, la Universidad de Alberta comenzó a utilizar la aplicación Archive-It para desarrollar este proyecto. Su primer año con Archive-It (2009) giró en torno a los sitios web heredados de la organización sin ánimo de lucro ya disuelta (según la correspondencia y conversación personal con Geoff Harder, junio de 2012).

A partir de 2010, la Universidad de Alberta comenzó a utilizar Archive-It como herramienta de desarrollo de colecciones más amplias. Los programas nacionales de archivado web no están tan desarrollados en Canadá como en otros países. Para suplir esta deficiencia, la biblioteca de la Universidad ha comenzado a recopilar en serio material de diversas áreas, como son, entre otras, las siguientes: temas políticos y económicos sobre las praderas canadienses, documentos gubernamentales, literatura gris o no convencional para las ciencias empresariales y de



EL CÍRCULO EXTERIOR - RECURSOS Y FLUJO DE TRABAJO

la salud, estudios circumpolares y materiales curriculares de educación provincial. De este modo, la visión de su programa Archive-It concuerda con su política de desarrollo de colecciones que no son de carácter digital. Dos de los grandes problemas que experimentan a la hora avanzar tienen que ver con el refinamiento de su estrategia de detección y la mejora de la visibilidad de sus colecciones. Les interesa especialmente determinar el modo más eficaz de facilitar el acceso a sus archivos web y a otras colecciones digitales. Como a la Universidad le interesan los estudios digitales, desean asegurarse de que los investigadores pueden utilizar sus colecciones de archivos web del mismo modo que ya utilizan otros recursos (correspondencia y conversación personal con Geoff Harder, junio de 2012).

La Biblioteca Estatal de Montana (MSL) constituye un ejemplo de visión institucional diferente. Su archivo web tiene por objeto archivar documentos estatales que actualmente solo se encuentran disponibles en línea. Su objetivo es «satisfacer las necesidades de información de los empleados de la agencia estatal, facilitar el acceso permanente a publicaciones estatales, ofrecer apoyo a las bibliotecas de Montana para que presten servicios y contenidos bibliotecarios de calidad, trabajar en la consoli-

ción de las bibliotecas públicas de Montana y facilitar el acceso a los recursos bibliotecarios a los ciudadanos de Montana con discapacidades físicas o visuales» (correspondencia personal con Beth Downs, James Kammer y Chris Stockwell, mayo de 2012). Un empleado de la Biblioteca Estatal de Montana resume del modo siguiente los motivos que tiene esta para archivar la web: «Dada la estrepitosa caída del número de publicaciones impresas y el auge exponencial inverso del número de publicaciones web, Archive-It ha reemplazado completamente la tradición histórica de la biblioteca depositaria estatal de adquirir y distribuir publicaciones impresas una a una» (correspondencia personal con Beth Downs, James Kammer y Chris Stockwell, mayo de 2012). Al comienzo de su suscripción en 2007, Montana estableció una política para administrar la mayoría de los aspectos de su programa de archivado web, incluidos los criterios de selección de contenidos que había que archivar, la frecuencia de rastreo y el alcance. Las interacciones entre Archive-It y MSL desde 2007 indican que este enfoque ha sido satisfactorio y está cumpliendo los objetivos de la biblioteca estatal.

1b. Recursos y flujo de trabajo

La fase de recursos y flujo de trabajo del ciclo de vida puede prestarse a varias interpretaciones. En el contexto del círculo exterior del modelo, las instituciones examinan los recursos y los flujos de trabajo que pueden ser aprovechados para crear o mantener el programa de archivado web de toda una institución. De este modo, los recursos y el flujo de trabajo pueden considerarse sinónimo de «política», ya que pueden aplicarse en diversos ámbitos del Modelo de Ciclo de Vida del Archivado Web. Los recursos y el flujo de trabajo también deberían considerarse términos generales de gestión del programa susceptibles de aplicarse a cada uno de los elementos del anillo interior del modelo. En este contexto, los recursos y el flujo de trabajo forman parte de las actividades cotidianas de archivado web. Por ejemplo, ¿cuánto tiempo puede dedicar una institución a examinar sus rastreos y cuántas personas deberían añadir sitios web a la aplicación Archive-It? En posteriores apartados del presente documento se tratarán

Personal de archivo	64%
Personal de biblioteca	42%
Personal de proyectos digitales	30%
Personal de tecnología de la información	8%
Otros (como estudiantes o «equipo web»)	8%

TABLA 1:
Tipos de empleados de una institución que trabajan con Archive-It

en profundidad los flujos de trabajo específicos de la gestión.

Uno de los recursos clave de los que disponen las organizaciones son sus empleados. Algunos de los intensos debates con varios socios de Archive-It durante la primera y el verano de 2012 y la encuesta realizada por la Universidad de Marquette revelan datos muy completos sobre los modelos de dotación de personal existentes en una amplia variedad de instituciones asociadas a Archive-It (Sweetser, 2011).

De las 37 instituciones que respondieron a la encuesta de Marquette, un tercio cuenta con dos o más personas dedicadas a Archive-It y más del 25% cuenta con cuatro o más personas. La encuesta también reveló que la mitad de las instituciones encuestadas dedican menos de una hora a la semana a trabajar con sus cuentas de Archive-It y el 44% dedica entre una y cinco horas a la semana a trabajar con la aplicación. La encuesta de Marquette también pedía a los encuestados que describieran los tipos de personas que trabajan en Archive-It. La Tabla 1 recoge los resultados; adviértase que los encuestados podían elegir más de un agrupamiento de personal, por lo que los resultados no suman un 100%.

Las conversaciones mantenidas con los seis socios de Archive-It citados en este documento revelaron resultados similares a los de la encuesta de Marquette. Los socios aportaron detalles de los empleados que trabajan con Archive-It, como son el número de

INSTITUCIÓN	NÚMERO DE EMPLEADOS PARTICIPANTES	DETALLES DE DOTACIÓN DEL PERSONAL
Universidad de Columbia	1 + participación de otros empleados	Actualmente (2012) un responsable de contenidos web realiza los rastreos, evalúa títulos y descripciones (<i>seeds</i>) y gestiona la cuenta de Archive-It, aunque han tenido dos de estos responsables en el pasado. Los estudiantes, los responsables de metadatos y el programador web también utilizan distintas partes de la aplicación de forma más limitada.
Universidad de Creighton	1	Creighton cuenta con un archivista a jornada completa y entre sus responsabilidades se encuentra administrar Archive-It; también le ayudan un poco otras personas de la biblioteca.
Universidad de Alberta	1 responsable técnico, con hasta 40 personas que se conectan activamente a la aplicación	Alberta cuenta con una gran red de personas que utilizan Archive-It activamente, muchas de las cuales son especialistas en determinadas materias.
Biblioteca Estatal de Montana	3	Los usuarios más activos son el bibliotecario de publicaciones estatales (que supervisa el programa), el catalogador de metadatos y el programador o analista de los sistemas bibliotecarios, que se ocupa de los problemas técnicos.
Archivos y Biblioteca Estatal de Carolina del Norte	4	La gestión de Archive-It se divide por igual entre dos representantes procedentes de la biblioteca estatal y los archivos estatales.

TABLA 2:
Número y tipos de empleados que trabajan con Archive-It

empleados y la naturaleza de su trabajo. La Tabla 2 recoge un resumen de los resultados, que presentan similitudes con los obtenidos en la encuesta de Marquette. La mayoría de los empleados suelen provenir de la biblioteca o los archivos (esta autora supone que los especialistas en materias concretas y los responsables de metadatos forman parte del personal bibliotecario), con la participación añadida del personal de tecnología de la información y los estudiantes.

Además de la cuestión de la dotación del personal, los recursos y el flujo de trabajo de este modelo comprenden también el modo en que las instituciones gestionan otros recursos. Por ejemplo, la Universidad de Columbia utiliza una base de datos interna para realizar un seguimiento de la información que no puede incorporarse a la aplicación Archive-It, como es la información administrativa y los datos de permisos de sitios con los que han contactado. Otro ejemplo es la decisión de colaborar y dividir la gestión del programa de archivado web entre la Biblioteca Estatal de Carolina del Norte y los Archivos Estatales de Carolina del Norte. Ambas instituciones gestionan una única colección de sitios web de agencias del Gobierno estatal. Al dividirse el trabajo diario, estas dos agencias cuentan con varios flujos de trabajo bien establecidos, que han desarrollado desde que comenzaron a utilizar Archive-It en 2005. La biblioteca y los archivos estatales se alternan la responsabilidad de realizar los rastreos y ambas instituciones llevan un control de calidad de los datos recopilados. Los empleados han ido rotando con los años; sin embargo, pese a esta rotación, las instituciones han llegado a la conclusión de que su asociación ha resultado ser una «colaboración fácil de mantener» (correspondencia personal con Kelly Eubank, Lisa Gregory, Kathleen Kenney y Rachel Trent, junio de 2012).

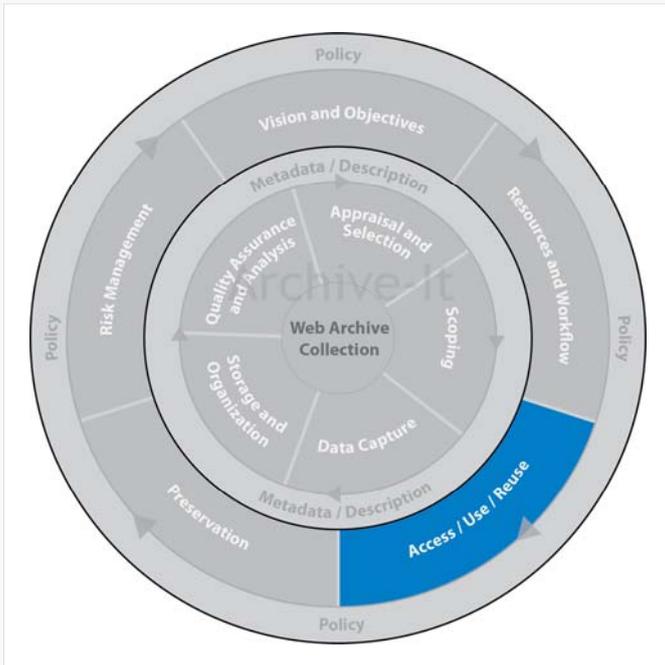
De las seis instituciones que utilizan Archive-It señaladas en este documento, la Universidad de Alberta es la que cuenta con el programa de archivado web de mayor envergadura en términos de dotación de personal. Alberta comenzó a usar Archive-It en 2009 con un reducido equipo de varias personas que ha ido creciendo hasta ser veintidós las que contri-

buyen activamente al programa. También han incorporado al trabajo a una serie de especialistas en determinadas materias. Además, el equipo cuenta con un bibliotecario de documentos gubernamentales y otro de metadatos que participan en la aplicación. Un representante del área de tecnología de la información ofrece apoyo a estas personas y filtra sus dudas para transmitir las al personal de Archive-It de Internet Archive. En un nivel superior, la biblioteca dispone de un «grupo de trabajo digital» compuesto por empleados de la biblioteca. Este grupo, formado principalmente por personas del área de desarrollo de colecciones, ayuda a configurar la política de archivado web en general y el uso de Archive-It en particular. Asimismo, un grupo de usuarios de Archive-It, con una amplia base de miembros, desarrolla y comparte conocimientos sobre Archive-It.

A diferencia de la Universidad de Alberta, la de Creighton cuenta con un único archivero que gestiona la suscripción a Archi-

ve-It de la Universidad y que también abogó por ello inicialmente como recurso necesario. David Crawford conoció Archive-It en la conferencia de la Society of American Archivists (Sociedad de Archivistas de Estados Unidos) de 2008 y trabajó para conseguir apoyos y establecer una suscripción a Archive-It en Creighton. Finalmente obtuvo la donación de un miembro de la junta para comenzar su programa de archivado web financiando una suscripción a Archive-It. Una herramienta como Archive-It permite a Crawford cumplir su objetivo de archivar la presencia web de la universidad, lo que no habría podido hacer por sí solo por falta de conocimientos especializados internos (conversación con David Crawford, julio de 2012). La experiencia de Crawford de tener que conseguir apoyo él solo para el archivado web parece coincidir con otras interacciones que Internet Archive ha tenido con algunas pequeñas instituciones como Creighton. Las instituciones pequeñas suelen tardar más en poner en marcha su programa por la carencia de personal y recursos económicos.

Las instituciones deben analizar cómo aprovechar sus recursos y dotaciones para crear o mantener el programa de archivado web



EL CÍRCULO EXTERIOR - ACCESO, UTILIZACIÓN Y REUTILIZACIÓN

Algunas universidades más pequeñas han formado consorcios para respaldar sus programas de archivo web y poder ampliar sus recursos compartidos (véase, por ejemplo, el triple consorcio de Bryn Mawr, Swarthmore y Haverford <http://www.archive-it.org/organizations/74>, uno de los socios piloto originales de Archive-It).

1c. Acceso, utilización y reutilización

Es fundamental establecer las políticas de acceso, utilización y reutilización para que el programa de archivo web tenga éxito. Las instituciones se plantean si permiten el acceso abierto a sus archivos web y el modo de hacerlo, si promocionan las colecciones y cómo hacerlo, además de cómo administrar el uso público del material. Gestionar estos procesos constituye el principal objetivo de la fase de acceso, utilización y reutilización del Ciclo de Vida del Archivado Web.

Parte de la elaboración de una política de acceso comprenderá la elección de la tecnología o herramienta específicas para facilitar el acceso a las páginas web archivadas. No obstante, a los efectos de este modelo, nosotros consideramos en cambio las

decisiones políticas de niveles superiores relacionadas con el acceso. Esto se debe, en parte, a que todos los individuos entrevistados para este proyecto acceden a los archivos web mediante el software Wayback, una herramienta de visualización de código abierto que permite al público examinar páginas web archivadas como si se tratara de páginas web activas.

La mayoría de los socios de Archive-It tienen sus contenidos archivados a disposición del público, aunque cada vez son más los que solicitan que se restrinjan determinados contenidos durante un tiempo: una URL específica, una colección particular o toda la cuenta con diversas colecciones. El equipo de Archive-It empieza a recibir más peticiones de restricción de contenidos por dirección IP para que las salas de lectura de las bibliotecas universitarias tengan más flexibilidad de acceso. (Nota: se puede disponer de esta función desde abril de 2013.)

Los socios de Archive-It pueden dirigir a los usuarios al sitio web de Archive-It para acceder a las colecciones (<http://www.archive-it.org>) o bien pueden vincular sus colecciones desde sus propias páginas mediante un cuadro de búsqueda o enlaces al software Wayback. Pueden aplicarse ambos enfoques a los socios, dependiendo de sus necesidades de acceso. Por ejemplo, la Biblioteca Estatal de Carolina del Norte y los Archivos Estatales de Carolina del Norte dan acceso a sus colecciones de Archive-It desde su propio portal web. Han elaborado un portal sólido que facilita información sobre los archivos web dirigida al público y a profesionales de la información, además de instrucciones para poder utilizar dichos archivos (<http://webarchives.ncdcr.gov/>). Por otra parte, la Universidad de Creighton ha adoptado un enfoque distinto. Dirigen a sus usuarios al sitio web de Archive-It para que accedan a las colecciones y no dan acceso a ellas desde su página web institucional. En palabras de David Crawford, prefieren que sus usuarios sean «autodirigidos» (conversación con David Crawford, julio de 2012).

Al igual que la Biblioteca Estatal de Carolina del Norte y los Archivos Estatales de Carolina del Norte, la Biblioteca Estatal de Montana creó también su pro-

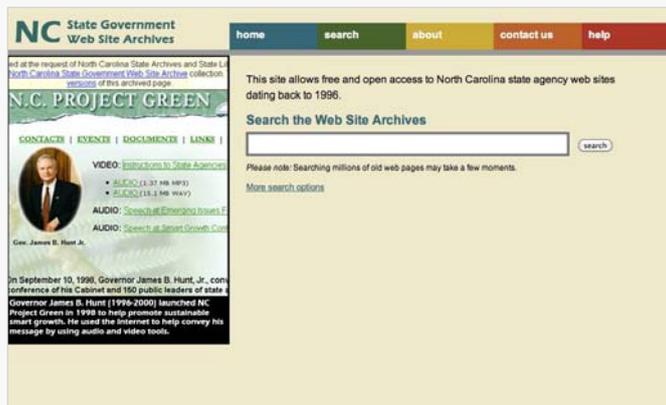


IMAGEN 1:
Página de inicio de los Archivos Web del Gobierno Estatal de Carolina del Norte, <http://webarchives.ncdcr.gov/>

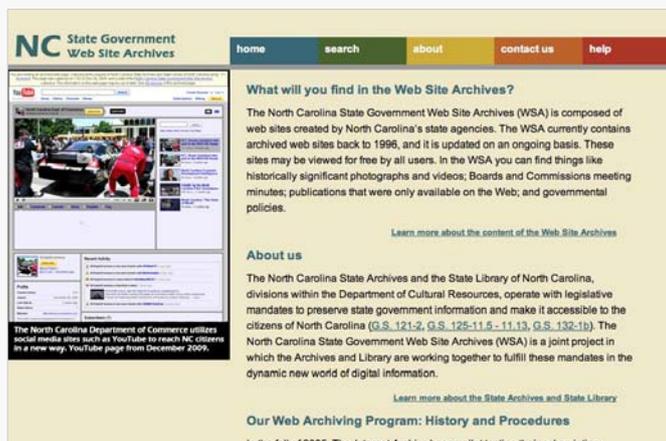


IMAGEN 2:
La sección «Acerca de» de los Archivos Web del Gobierno Estatal de Carolina del Norte, <http://webarchives.ncdcr.gov/about.html>

pio portal web que da acceso a sus colecciones de Archive-It (http://msl.mt.gov/For_State_Employees/connect/default.asp). Además de facilitar el acceso a los datos recopilados mediante el servicio de Archive-It, la Biblioteca Estatal de Montana extrajo páginas web más antiguas que se remontan a 1996 del archivo web general de Internet Archive. Se puede acceder a estas páginas web desde el portal junto con los datos de Archive-It, que se remontan a 2006. El objetivo de la biblioteca para dar acceso a través de su propio sitio web es «crear una única marca identificable asociada a la información del Gobierno estatal» (correspondencia personal con Beth Downs, James Kammerer y Chris Stockwell, mayo de 2012). Montana también ha encontrado otros modos inno-

vadores de dirigir la atención hacia sus archivos web. Al pie de todas las páginas web de la Biblioteca Estatal de Montana se encuentra el enlace «historial de la página». Estos enlaces dirigen a los visitantes a las versiones archivadas de la página web, de modo que pueden ver cómo ha cambiado con el tiempo. Por ejemplo, el «historial de la página» que figura en la página inicial de la biblioteca estatal <http://msl.mt.gov/>¹ dirige al visitante a una lista de fechas de capturas de fácil acceso de esta página web: <http://wayback.archive-it.org/499/query?type=urlquery&url=http://msl.mt.gov/&dates=>.



IMAGEN 3:
Página inicial de la Biblioteca Estatal de Montana (<http://msl.mt.gov/>)

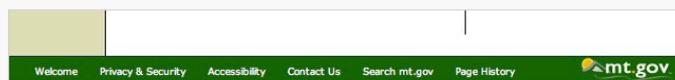
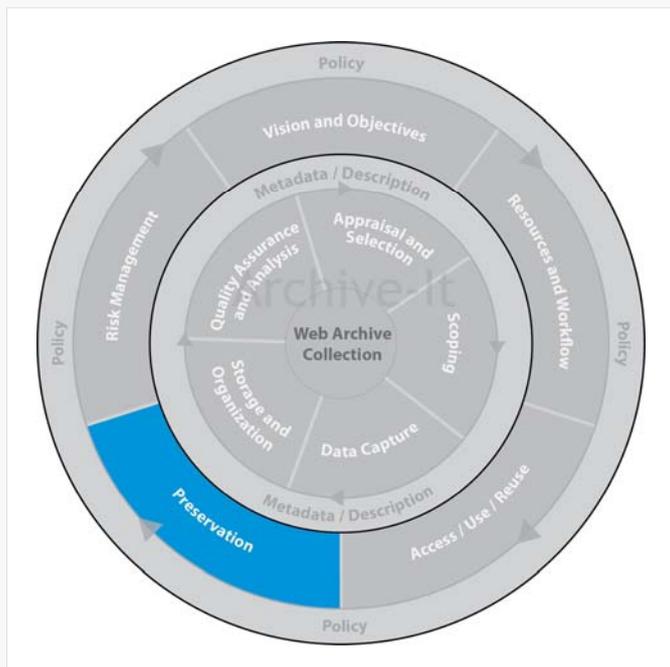


IMAGEN 4:
Detalle del pie de la página inicial de la Biblioteca Estatal de Montana



EL CÍRCULO EXTERIOR - CONSERVACIÓN

1d. Conservación

Los datos recabados para la elaboración de este documento sugieren que la conservación es una cuestión en desarrollo para las instituciones que archivan la web que va de la mano de la naturaleza evolutiva de la conservación digital y el desarrollo de repositorios digitales. El equipo de Archive-It detectó que sus socios tendían a emplear distintas estrategias de conservación. Muchas instituciones que trabajan con el servicio Archive-It recurren a Internet Archive para almacenar y conservar sus ficheros WARC y los metadatos asociados a los mismos. Hay varios socios que también transfieren sus datos a un disco duro local o descargan sus ficheros WARC directamente de los servidores de Internet Archive. Algunas instituciones asociadas están tratando de incorporar los ficheros WARC a su repositorio digital local, aunque estos proyectos aún están en una fase muy precoz.

Según una encuesta reciente realizada por los socios de Archive-It, estos desean conservar sus datos y tener múltiples copias de los mismos en diversos lugares. Sin embargo, están teniendo dificultades para conseguirlo. En la encuesta, el 56% de los encuestados respondió que le gustaría archivar sus da-

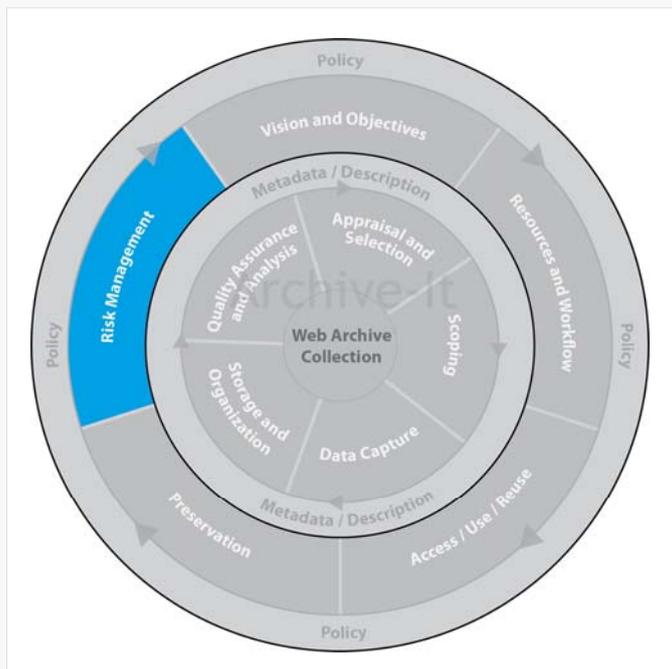
tos en su propio repositorio local (independientemente de la plataforma que utilicen).

Sin embargo, el 31% de los socios indicaron que prefieren almacenar sus datos en Internet Archive porque están satisfechos con esa estrategia o porque no disponen de medios para conservar los datos en otro lugar. Aproximadamente el 60% de los encuestados no dispone aún de un repositorio digital local. Las dos razones esgrimidas con mayor frecuencia por las que no disponían de un repositorio son que «no estamos seguros de nuestras necesidades» y «estamos sopesando qué sistema elegir». Estos resultados, junto con la información anecdótica obtenida con los años de los socios de Archive-It, sugieren inequívocamente que los socios están teniendo dificultades con el modo de conservar los datos que recopilan mediante el archivado web, por lo que cabe esperar que se produzcan importantes avances en este ámbito del modelo en los próximos años.

1e. Gestión del riesgo

A la hora de desarrollar un programa de archivado web, muchas instituciones consideran el nivel de riesgo relacionado con los derechos de autor que están dispuestos a aceptar y la manera en que gestionarán dicho riesgo. Uno de los ejemplos más claros de formulación activa de políticas de gestión del riesgo es si las instituciones solicitan permiso a los propietarios de sitios web antes de archivarlo y la manera en que lo hacen. El servicio de Archive-It utiliza desde hace tiempo robots.txt (una sencilla solución tecnológica) como herramienta de gestión de permisos, ya que permite a los propietarios de sitios web excluir automáticamente sus sitios del proceso de archivado. Además del protocolo robots.txt, los socios de Archive-It a veces solicitan a los propietarios de sitios web un permiso por escrito antes de comenzar a recopilar.

Por ejemplo, la Universidad de Columbia se pone en contacto directamente con los propietarios de sitios y les solicita formalmente un permiso para archivar sus sitios web antes de comenzar a recopilarlos. Se trata de un proceso de varias semanas de duración



EL CÍRCULO EXTERIOR - GESTIÓN DEL RIESGO

en el que se contacta dos veces con el propietario del sitio. Si no se obtiene respuesta al primer contacto al cabo de tres semanas, el equipo de la Universidad de Columbia envía un mensaje de recordatorio. Si continúan sin recibir respuesta al cabo de otras tres semanas, proceden a realizar la recopilación. En general, el índice de respuestas de la Universidad de Columbia es del 52%: de los 783 sitios contactados, 400 respondieron y concedieron su permiso, 378 no respondieron y solo 5 propietarios de sitios web respondieron negativamente solicitando que no se archivaran sus sitios (correspondencia personal con Alex Thurman, febrero de 2013). Del mismo modo, la Universidad de Alberta solicita permiso de manera selectiva a los sitios web que archivan. Esta decisión se tomó tras consultar con su departamento jurídico, que les dio un «umbral de riesgo» que debían seguir, por lo que solicitan permiso cuando consideran que deben respetar este umbral (correspondencia y conversación personal con Geoff Harder, 27 de junio de 2012).

Las decisiones de gestión del riesgo también resultan evidentes en las decisiones que toman las instituciones a la hora de elegir qué sitios van a archivar. Originalmente, la Biblioteca Estatal de Carolina del

Norte y los Archivos Estatales de Carolina del Norte solo recopilaban los sitios web de las agencias estatales. Sin embargo, en 2009 comenzaron a recopilar las fuentes de noticias (*feeds*) de las agencias estatales en páginas web de redes sociales como Facebook, Twitter y Flickr. Pese a que el contenido se hallaba en un sitio web de terceros y no era controlado por una agencia estatal de Carolina del Norte, los archivistas y bibliotecarios decidieron dar un paso adelante en el archivado tras sopesar los posibles riesgos y resultados (correspondencia personal con Kelly Eubank, Lisa Gregory, Kathleen Kenney y Rachel Trent, junio de 2012).

No todas las organizaciones solicitan permiso antes de recopilar contenidos y muchas otras tienen claro que, como archivo o biblioteca, sus organizaciones tienen el derecho y la obligación de recabar contenidos accesibles al público en la web activa. «Uso razonable» y «juego limpio» son dos términos que el equipo de Archive-It escucha de sus socios a la hora de decidir si recopilan contenidos web accesibles al público. En muchos casos, el mandato de una organización llega a ignorar los robots.txt en las hojas de estilo en cascada (CSS), por lo que la página web archivada se recopila completamente. En otros casos, esta política contempla que los investigadores e historiadores puedan recopilar documentos o sitios web con el fin de poder presentar una instantánea precisa y completa sobre un tema y cada vez más se incluyen contenidos de acceso público de las redes sociales.

El riesgo puede gestionarse y mitigarse de manera preventiva; por tanto, hay ocasiones en que las instituciones también deben resolver problemas que pueden surgir después de archivar el contenido. En la Universidad de Creighton, a un fotógrafo le molestó que su sitio web hubiera sido archivado, pese a que formaba parte del espacio web más amplio de la Universidad y, por tanto, fue rastreado de acuerdo con la política de gestión de documentos de esta. Creighton decidió eliminar la página web del archivo y trabajó con el equipo de Archive-It para resolver este problema, por lo que al cabo de unas horas se retiró el contenido. Desde entonces, Creighton ha decidido que, si existe el riesgo de incomodidad o

litigio, se retirará el contenido del archivo web (conversación con David Crawford, julio de 2012).

Nota: El servicio de Archive-It no adopta postura alguna sobre los derechos de autor y respeta la política de Oakland en materia de archivo (*Oakland Archive Policy*), adoptada en 2002, en un intento por trabajar de manera colaborativa con los proveedores de contenidos. El servicio acepta las solicitudes de retirar contenidos para que no sean accesibles al público.

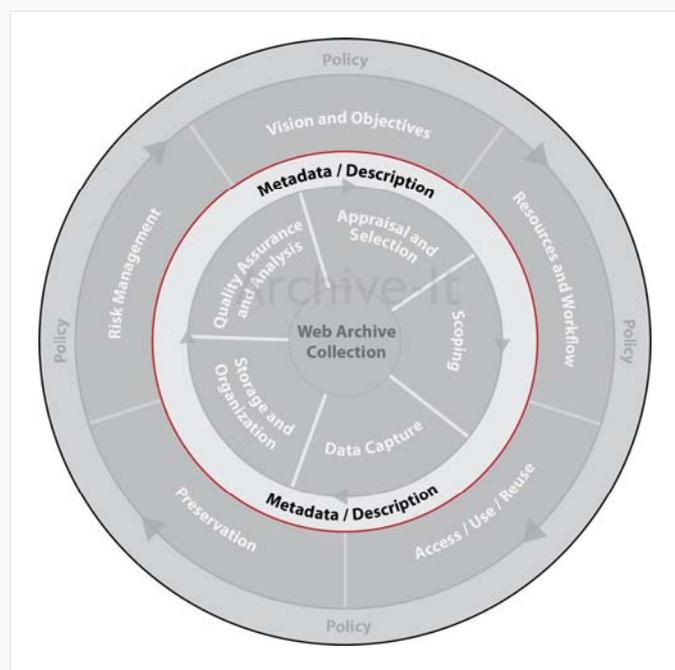
2. Banda gris

2a. Metadatos y descripción

A partir de la información de sus socios, el equipo de Archive-It concluyó que la parte relativa a metadatos y descripción del ciclo de archivado web, al igual que la de política, se solapa considerablemente con otros pasos del ciclo. Por tanto, se tomó la decisión de presentar los metadatos y la descripción como una banda global del modelo en lugar de una parte en sí misma independiente del proceso. Al igual que ocurre con la mayoría de los aspectos del archivado web, las prácticas más recomendadas están evolucionando en cuanto al uso y la creación de metadatos y tendencias descriptivas para los archivos web. No obstante, el equipo de Archive-It puede extraer conclusiones basadas en el modo en que las instituciones utilizan la función de metadatos y descripción en Archive-It. Los datos recopilados internamente por el equipo de Archive-It en 2011 indican que más del 70% de los socios de Archive-It genera metadatos a nivel de recopilaciones, más del 60% genera metadatos de título y descripción (*seed*) y el 10% genera metadatos a nivel de documentos. Los títulos y descripciones constituyen el punto de partida de las URL para los rastreos web y los documentos son las páginas web archivadas de forma individual. Además, estos mismos datos revelaron que el 60% de los socios crea metadatos de recopilaciones y títulos y descripciones. Algunos socios, como la Universidad de Columbia, generan una importante cantidad de metadatos para sus colecciones de Archive-

It y trabajan con Archive-It para cambiar y ampliar la funcionalidad de los metadatos de la aplicación. Aunque no se dispone de datos estadísticos anteriores sobre la generación de metadatos, basándose en pruebas anecdóticas, el equipo de Archive-It cree que ha aumentado el ritmo al que los socios crean metadatos. La encuesta de Marquette constata estas conclusiones. En la encuesta se preguntaba cómo utilizaban los socios de Archive-It las funciones descriptivas de la aplicación. Entre las principales conclusiones de la encuesta se encuentran las siguientes:

- El 35% de los encuestados prepara metadatos a nivel de colecciones más allá del campo de descripción solicitado; otro 35% no lo hace.
- El 81% de los encuestados no prepara metadatos para documentos individuales recopilados mediante rastreos de Archive-It.



BANDA GRIS - METADATOS Y DESCRIPCIÓN



CÍRCULO INTERIOR: VALORACIÓN Y SELECCIÓN

- El 75% de los que preparan metadatos para documentos individuales los generan manualmente en vez de recabarlos del sitio web.
- La mayoría de los encuestados no catalogan contenidos de Archive-It a ningún nivel dentro de un registro de catálogos (colección, semilla [seed], documento) (Sweetser, 2011).

En general, los autores de la encuesta de Marquette creen que uno de los principales resultados de su trabajo sugiere que los socios de Archive-It no generan metadatos de sus colecciones en la propia aplicación Archive-It. Sweetser propone tres razones posibles: «Las organizaciones aún no se han puesto a elaborar metadatos en Archive-It y siguen en una fase embrionaria en lo que respecta a sus esfuerzos de archivado web. Las organizaciones no creen que los metadatos estén justificados o que su creación resulte de utilidad [y] se están centrando en sus prácticas de creación de metadatos en áreas ajenas a la plataforma Archive-It» (Sweetser, 2011).

3. El círculo interior

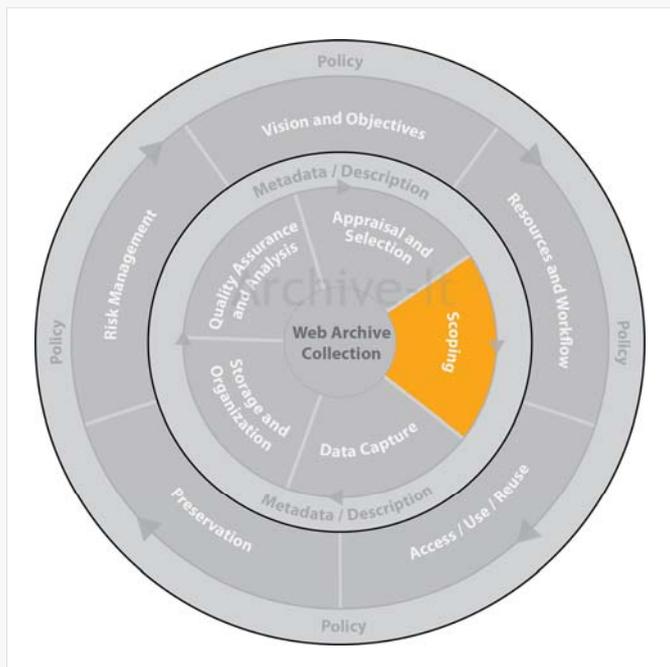
Las fases anteriores del ciclo de vida forman parte del ciclo exterior del modelo, que está relacionado con las preguntas más amplias acerca de la creación y definición de un programa de archivado web institucional. Las fases restantes del modelo, las del círculo interior, describen las actividades cotidianas propias de la gestión de un programa de archivado web.

3a. Valoración y selección

La fase de valoración y selección del archivado web entraña elegir sitios web específicos para recopilar. Este paso implica más elementos granulares y específicos de decisión que la fase más amplia de la política sobre «visión y objetivos» del ciclo de vida. A la hora de crear una política, las instituciones conciben planes generales para todo el programa, como qué materias se incluirán en las actividades de recopilación. Sin embargo, en la fase de valoración y selección las instituciones eligen las URL específicas que van a archivar. Tal y como indican los ejemplos comentados a continuación, estas elecciones pueden realizarse de diversas maneras, con la aportación de distintos tipos de personas.

Los archivos y bibliotecas estatales, por ejemplo, suelen centrar sus esfuerzos de archivado web exclusivamente en sitios web de agencias estatales y recopilan esas URL. Así sucede en la Biblioteca Estatal de Montana, la Biblioteca Estatal de Carolina del Norte y los Archivos Estatales de Carolina del Norte. No obstante, en el caso de Carolina del Norte, también archiva fuentes de noticias (*feeds*) de redes sociales que generan las agencias estatales en Facebook, Twitter y Flickr, porque consideran que estas noticias son ampliaciones de los registros web oficiales. Esta decisión política se describe con más detalle en la sección dedicada a la gestión del riesgo de este documento.

En ocasiones, las universidades que archivan la web enfocan la valoración de sitios web de manera diferente. Tienden a archivar la presencia de la web uni-



CÍRCULO INTERIOR: DEFINICIÓN DEL ALCANCE

versitaria o a crear colecciones basadas en temas específicos. Por ejemplo, los ámbitos temáticos principales de las recopilaciones de archivos web de la Universidad de Columbia y la Universidad de Alberta comprenden temas sobre los derechos humanos y la industria y cultura canadienses, respectivamente. Trasladar los objetivos principales de la institución a una lista de sitios que rastrear constituye el objetivo del proceso de valoración y selección. Para ello, la Universidad de Alberta, por ejemplo, trabaja con bibliotecarios de enlace de materias para elegir las URL. La fase de valoración y selección constituye un área en evolución de la que esperamos aprender gracias a nuestros socios.

3b. Definición del alcance

Tras elegir qué sitios se van a archivar, las instituciones deben decidir si desean archivarlos completa o parcialmente. Esta labor puede realizarse antes de capturar la primera página o después de recopilar el contenido, como parte de una revisión general de la calidad de la recopilación. Esta parte del ciclo de vida puede ser bastante técnica en función de las herramientas que utilice la institución.

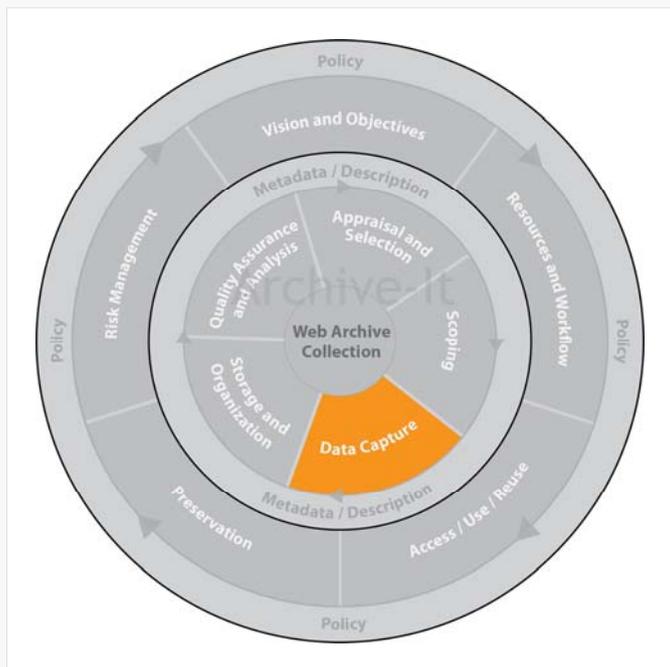
El servicio Archive-It ofrece a las instituciones varias herramientas para ajustar el alcance de sus rastreos. En primer lugar, los socios pueden limitar lo que desean rastrear elaborando un listado de solo una parte del sitio web como punto de partida del rastreo en vez de la totalidad del sitio web. Por ejemplo, una institución podría querer archivar <http://www.ncgov.com/government/index.aspx> en lugar de archivar <http://www.ncgov.com> y solo capturaría páginas pertenecientes a esa URL. Archive-It también incorpora otras herramientas para limitar en qué grado se rastrea un sitio. Los resultados de una encuesta reciente revelan que el 73% de los encuestados utilizan, al menos ocasionalmente, una herramienta de restricción de *hosts*. Esta herramienta permite a los socios bloquear *hosts* específicos o subsecciones de un sitio, para que no se archiven. Por ejemplo, una institución podría no desear recopilar imágenes de terceros que puedan estar alojadas en un sitio web específico. Restringir la duración de un rastreo mediante limitaciones de tiempo constituye la segunda herramienta más utilizada, según indicó el 64% de los encuestados.

Actualmente el 27% de los socios de Archive-It realiza rastreos para capturar solo documentos PDF y cabe esperar un

aumento de este porcentaje a medida que el formato PDF se extienda más por la web y sea cada vez más el único registro existente. El servicio Archive-It está investigado si conviene añadir esta opción para otros tipos de formatos de fichero. A medida que las redes sociales se convierten cada vez más en componentes esenciales para las actividades de recopilación, el servicio está estudiando maneras singulares de ofrecer una solución de captura y acceso a los medios sociales, principalmente Facebook, Twitter y YouTube, desde diciembre de 2012.

Como ya se ha comentado, el proceso de definición del alcance puede ser muy técnico. El equipo de la Universidad de Alberta se sorprendió ante las com-

Tras elegir qué sitios se van a archivar, las instituciones deben decidir si desean archivarlos completa o parcialmente



CÍRCULO INTERIOR: CAPTURA DE DATOS

plejidades que entraña definir el alcance de los rastreos de un modo eficaz. Han comprobado que deben reajustar sus políticas a medida que rastrean, adaptándose a veces al tipo de datos que realmente pueden recabar (correspondencia y conversación personal con Geoff Harder, 2012). De igual modo, Creighton también ha comprobado que definir el alcance de un rastreo entraña más trabajo; David Crawford a menudo necesita instruir a las personas del campus acerca del espacio web y procura indicar a los programadores web que tengan en cuenta las necesidades de rastreo a la hora de introducir cambios futuros en las páginas web (conversación con David Crawford, julio de 2012).

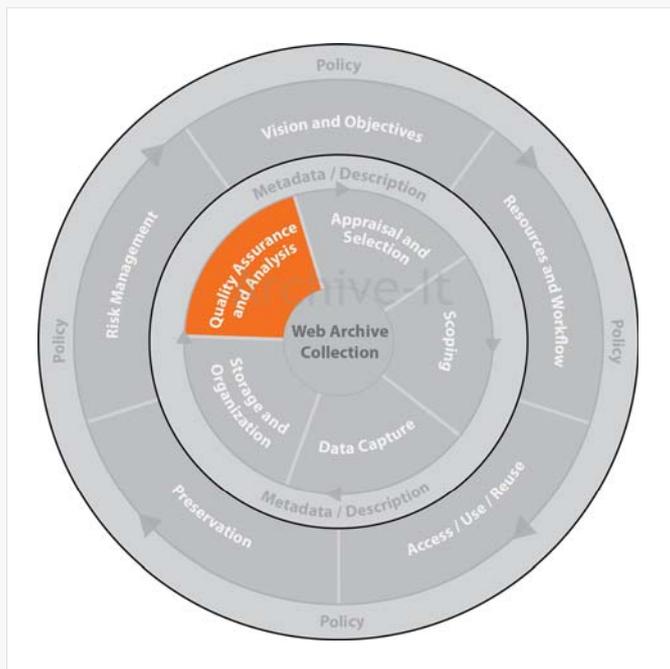
3c. Captura de datos

Una vez que las instituciones han decidido los sitios web que van a capturar y el modo de hacerlo, ponen en práctica sus planes en la fase de captura de datos del proceso. Aquí se tratarán los aspectos prácticos del software de rastreo. Determinarán la frecuencia y los tiempos de rastreo y cuándo han de interrumpir rastreos prolongados para, a continuación, comenzar con los rastreos. La aplicación Archive-It incorpora funciones que permiten a los socios ajustar la fre-

cuencia y duración en el rastreador web de código abierto (Heritrix).

Programar los rastreos para la captura de datos continua y reiterada constituye un ámbito en el que las instituciones que utilizan Archive-It ejercen un gran control sobre sus rastreos. Los datos recabados en 2011 indicaron que el 78% de todos los socios de Archive-It emplean más de una frecuencia de rastreo. Es decir, no rastrean todos sus sitios en un único intervalo, utilizan calendarios diferentes para diferentes recopilaciones y sitios web. En el momento en que se recabaron estos datos, las frecuencias de rastreo más extendidas fueron una vez, mensual y trimestral.

Dada la diversidad de los sitios web en cuanto a estructura y construcción, la fase de captura de datos del archivado web puede dar algunas sorpresas. Por ejemplo, un sitio puede ser mucho mayor de lo previsto y, por tanto, agotar los recursos de almacenamiento. De igual modo, los administradores de páginas web tienen formas de impedir que se archiven, lo cual puede exigir una intervención tecnológica o una negociación entre las partes implicadas. Por ejemplo, David Crawford, de la Universidad de Creighton, tuvo problemas a la hora de archivar sitios web que habían sido bloqueados por sus administradores web y, al hablar con ellos, se sorprendió de lo poco que sabían del funcionamiento interno de sus webs (correspondencia y conversación personal con David Crawford, julio de 2012). Para evitar sorpresas en la captura de datos, Archive-It permite a los socios utilizar una función de rastreo de prueba que genera informes sobre los datos capturados sin necesidad de capturarlos realmente. Esta opción permite a las instituciones ver lo que habrían archivado sin tener que utilizar sus recursos innecesariamente. Una reciente encuesta a socios de Archive-It reveló que el 69% de los encuestados realizan rastreos de prueba siempre o a menudo cuando van a añadir semillas nuevas o van a comenzar una nueva recopilación.



CÍRCULO INTERIOR: ASEGURAMIENTO Y ANÁLISIS DE LA CALIDAD

3d. Aseguramiento y análisis de la calidad

Una vez que las instituciones han capturado los datos de los sitios web, deben examinar lo que han archivado y evaluar su calidad e integridad. Esta tarea puede realizarse mediante informes generados por los rastreadores o examinando los propios archivos mediante una herramienta de acceso como el software Wayback. El proceso de archivado web puede comprender labores de prueba y error. Como sucede con la mayoría de los aspectos del archivado web, no ha surgido ni una sola práctica recomendada para el aseguramiento de la calidad entre las instituciones que archivan la web. No obstante, existen algunas tendencias comunes entre los socios de Archive-It en cuanto a los tipos de información de rastreo que examinan.

Los datos de la encuesta realizada por Archive-It revelan que la mayoría de los socios examinan a menudo o siempre los informes generados tras el rastreo como parte del servicio. Esto se debe a que las instituciones tienden a interesarse por cuánto material y exactamente de qué tipo están recopilando cuando comienzan un programa de archivado web. Las conclusiones que se extraen de la encuesta reali-

zada durante el verano de 2012 a los socios de Archive-It indican que el 68% de las instituciones encuestadas examinaron periódicamente los informes de *hosts*. Solo el 11% lo hace rara vez o nunca. Examinar los informes puede llevar tiempo y los analizados deben conocer qué anomalías deben buscar. Tres de los encuestados afirmaron que la falta de personal o de recursos dificulta el análisis de los informes después de cada rastreo. En 2011 el servicio implantó una herramienta de QA (aseguramiento de la calidad) y la posibilidad de realizar un rastreo de parcheo en URL de nivel superior que no habían sido capturadas completa-

mente la primera vez. La respuesta ha sido positiva y el servicio sigue trabajando en la ampliación de opciones de la herramienta de QA. En el momento de redactar este documento, poco se sabe de cómo los socios de Archive-It llevan a cabo exactamente el aseguramiento de la calidad de sus rastreos; y uno de nuestros objetivos es aprender más en este terreno a medida que las necesidades de los socios se hacen más palpables.

Una vez que las instituciones han capturado los datos de los sitios web, deben examinar lo que han archivado y evaluar su calidad e integridad

CONCLUSIONES Y PASOS SIGUIENTES

El Modelo de Ciclo de Vida del Archivado Web supone un paso más hacia la creación de un conjunto de prácticas recomendadas para crear y mantener un programa de archivado web. Tras más de siete años prestando servicio y trabajando con socios con visión de futuro, el equipo de Archive-It tiene claro que la web ya no es «una maraña» y que, en beneficio de todos, debemos seguir trabajando juntos para hallar soluciones de captura y presentación de contenidos web. A medida que la tecnología evoluciona y aumenta la información que se publica exclusivamente en línea, más instituciones de todo tipo nece-

sitarán archivar contenidos web. Muchos de los socios de Archive-It han sido pioneros en el archivado web y se complacen en compartir lo aprendido. Y aunque compartimos nuestros conocimientos en este documento, sabemos que la web y las mejoras

prácticas de archivado web continuarán evolucionando. El equipo de Archive-It estima que este modelo y las instituciones que lo aplican son lo bastante flexibles para crecer y evolucionar codo con codo con la web que están intentando archivar.

NOTAS

¹ Debido a las inminentes migraciones de plataforma, las URL de la Biblioteca Estatal de Montana podrán variar en un futuro próximo.

CONVERSACIONES O CORREO ELECTRÓNICO

Universidad de Alberta: conversación con Geoff Harder, 27 de junio de 2012.

Biblioteca Estatal de Montana: correspondencia con Beth Downs, James Kammerer y Chris Stockwell, 29 de mayo de 2012.

Biblioteca y Archivos Estatales de Carolina del Norte: correspondencia con Kelly Eubank, Lisa Gregory, Kathleen Kenney y Rachel Trent, 8 de junio de 2012.

Universidad de Creighton: conversación con David Crawford, 17 de julio de 2012.

Universidad de Columbia: correspondencia con Alex Thurman y Tessa Fallon, 17 de mayo de 2012 y 21 de febrero de 2013.

Datos estadísticos de AIT de 2011: correspondencia con Kristine Hanna, septiembre de 2012.

OBRAS CITADAS

Sweetser, Michelle (2011). *Archive-It Metadata Usage Survey Results* [diapositivas en PowerPoint]. Fuente: <https://webarchive.jira.com/wiki/display/ARIH/Archive-It+Meeting+Presentations+2011>