

Los Caminos de los Documentos de Archivo Digitales: Tópicos en Preservación Digital

Módulo 7: Gestión y Preservación de Documentos de Archivo en Ambientes Web

InterPARES / ICA
Primera Versión en Español
Octubre, 2013
Traducción del inglés a cargo de:
Alicia Barnard, Alejandro Delgado y Juan Voutssás

Tabla de Contenido

Los Caminos de los Documentos de Archivo: Tópicos en Preservación Digital

1	Prefacio	3
---	----------------	---

1.1	Acerca del ICA e InterPARES.....	4
1.2	Público Objetivo	6
1.3	Como Usar los Módulos	6
1.4	Objetivos	6
1.5	Alcance	7
1.6	Base de Datos Internacional de Terminología	8
Módulo 7: Gestión y Preservación de Documentos de Archivo en Ambientes Web		3
2	Introducción	8
2.1	Propósito y Objetivos.....	8
2.2	Resultados del Aprendizaje.....	9
2.3	Terminología.....	9
3	Estrategia de Preservación de un Sitio Web	10
3.1	Capacidades Tecnológicas	10
3.2	Política / Requisitos de Gestión Archivística.....	11
3.2.1	<i>Gestión Archivística</i>	12
3.3	Metadatos.....	12
3.4	Gestión de Derechos/ Derechos de Propiedad Intelectual	13
3.5	Desarrollo y Capacitación del Personal	14
3.6	Descripción de Recursos, Documentación y Acceso.....	15
3.7	Planeación para la Recuperación en Casos de Desastre.....	15
3.8	Verificaciones de Validación.....	16
3.9	Formatos de Archivo.....	16
3.10	Medios de Almacenamiento	19
3.11	Normas.....	22
3.12	Mantenimiento de Documentos de Archivo basados en la Web en el Tiempo.....	23
3.13	Métodos / Herramientas para captura en Sitios Web	24
3.13.1	<i>Transferencia Directa</i>	24
3.13.2	<i>Cosecha Remota</i>	25
3.13.3	<i>Sitio Web Espejo</i>	26
3.13.4	<i>Herramientas para la Captura de un Sitio Web</i>	26
4	Plan General de Acción para la Preservación de un Sitio Web.....	29
5	Estudio de Caso: Desarrollo de un Plan para la Preservación de un Sitio Web para una Asociación Estudiantil en una Institución Académica	31
5.1	Antecedentes sobre la Organización	31
5.2	Los Retos	32
5.3	El Proceso de Desarrollo.....	32
6	Preguntas de Repaso	34
7	Recursos Adicionales	35

Los Caminos de los Documentos de Archivo Digitales: Tópicos en Preservación Digital

1 Prefacio

Los Caminos de los Documentos de Archivo Digitales : Tópicos en Preservación Digital Archivística es una iniciativa educacional desarrollada conjuntamente por el Consejo Internacional de Archivos –ICA- y el proyecto de “Investigación Internacional en Preservación de Documentos de Archivo Permanentes en Sistemas Electrónicos” - International Research on Permanent Authentic Records in Electronic Systems- o proyecto InterPARES. Su propósito es ofrecer capacitación a archivistas y profesionales de los documentos de archivo en la producción, gestión y preservación de documentos de archivo digitales auténticos, fiables y usables. El programa presupone que el usuario de estos módulos cuenta con una sólida base en los conceptos básicos de la administración de documentos de archivo y en la teoría archivística, y sienta sus bases en ese conocimiento.

“*Los Caminos de los Documentos de Archivo Digitales*” está conformada por ocho módulos se orienta al conocimiento teórico y práctico necesario para establecer el marco de referencia, la estructura de gobernanza y los sistemas requeridos para manejar y preservar documentos de archivo digitales a través de su “ciclo de vida”. Cada módulo trata específicamente un tema relevante para la administración o la preservación de los documentos de archivo. El programa completo se ofrece sin costo en el sitio Web del ICA en: <http://www.ica.org>

1.1 Acerca del ICA e InterPARES

El ICA e InterPARES tienen el compromiso de crear materiales educativos para la educación continua de archivistas y administradores de documentos de archivo, construir conocimiento básico, diseminar los nuevos hallazgos y dotar a los archivistas y profesionales de los documentos de archivo con el conocimiento y las competencias especializados necesarios para manejar y preservar de documentos de archivo digitales.

El Consejo Internacional de Archivos (ICA) (www.ica.org) está dedicado al manejo eficaz y a la preservación de documentos de archivo, así como al cuidado y uso del patrimonio archivístico mundial a través de su representación de profesionales de la gestión documental y archivos a lo largo del planeta. Los archivos son un recurso inmenso: son un derivado o sustancia documental del quehacer humano y como tales testigos irremplazables de eventos pasados, fundamento de la democracia, de la identidad de individuos y comunidades, así como de los derechos humanos. Pero también son frágiles y vulnerables. El ICA se esfuerza por proteger y asegurar el acceso a los archivos mediante asesorías, el establecimiento de estándares, el desarrollo profesional y el impulso del diálogo entre archivistas, tomadores de decisiones, productores y usuarios de archivos.

El ICA es una organización neutral no gubernamental, fundada por sus miembros, los cuales operan por medio de las actividades propias de cada membresía. Por más de sesenta años el ICA ha unido a instituciones archivísticas y profesionales lo largo del mundo para asesorar acerca de la buena administración archivística y la protección física del patrimonio así registrado, para producir estándares reconocidos y buenas prácticas e impulsar el diálogo, el intercambio y la disseminación de ese conocimiento y experiencia más allá de fronteras nacionales. Con aproximadamente 1500 miembros en 195 países y territorios, los valores del Consejo han sido aprovechar la diversidad cultural de sus integrantes para entregar soluciones eficaces y una profesión flexible e imaginativa.

El Proyecto InterPARES - The International Research on Permanent Authentic Records in Electronic Systems o Proyecto de Investigación Internacional en Preservación de Documentos de Archivo Permanentes en Sistemas Electrónicos - (www.interpares.org) pretende desarrollar conocimiento original esencial para la preservación a largo plazo de documentos de archivo producidos y conservados en formatos digitales, así como proporcionar una base sólida para estándares, políticas, estrategias y planes de acción para asegurar la longevidad de esos materiales documentales y la capacidad de sus usuarios para poder confiar en su autenticidad. El proyecto InterPARES se ha desarrollado en tres etapas:

InterPARES 1 (1999-2001) el cual se enfocó en el desarrollo de teoría y métodos para asegurar la preservación de la autenticidad de los documentos de archivo producidos y/o mantenidos en bases de datos y sistemas de gestión archivística de archivo en el transcurso de actividades administrativas. Los hallazgos de esta etapa presentaron el punto de vista del preservador de los documentos de archivo.

InterPARES 2 (2002-2007) continuó investigando acerca de temas relativos a la autenticidad, y se examinaron temas sobre fiabilidad y exactitud durante todo el ciclo de vida de los documentos de archivo, desde su producción hasta su preservación permanente. Se enfocó en aquellos documentos de archivo producidos en entornos digitales dinámicos e interactivos a lo largo de actividades artísticas, científicas y gubernamentales.

InterPARES 3 (2007-2012) se construyó con base en los hallazgos 1 de InterPARES y 2, así como en proyectos de preservación digital de otras partes del mundo. Llevó la teoría a la práctica trabajando con archivos y unidades archivísticas dentro de organizaciones con recursos humanos y/o financieros limitados con el fin de implementar en ellas programas sólidos de gestión y preservación de documentos de archivo.

1.2 Público Objetivo

El público objetivo para el cual está destinado este programa consiste en archivistas y profesionales de la gestión archivística y la información interesados en ampliar sus competencias en la administración de documentos de archivo digitales. En conjunto, los módulos conforman todo un paquete de recursos documentales para la educación continua de profesionales, con especial énfasis en aquellos temas que influyen en la preservación de documentos de archivo auténticos, fiables y exactos.

1.3 Cómo Usar los Módulos

Cada módulo está conformado por conocimiento teórico y metodológico, y sus aplicaciones prácticas ilustradas a través de casos de estudio y escenarios modelo. Si bien los módulos fueron desarrollados por el *Team* Canadá de InterPARES y por tanto ejemplificados con un contexto canadiense, cada uno es adaptable a un dominio específico o contexto jurídico. Para una mayor aplicabilidad, han sido traducidos a los idiomas de los socios del ICA.

Los módulos pueden ser estudiados individualmente o como un conjunto de acuerdo con una necesidad o interés, mismos que cubren el rango de competencias requeridas. Pueden ser auto-estudiados por personas u ofrecidos por medio de asociaciones profesionales o capacitación laboral. Los módulos también contienen plantillas que permiten a universidades o asociaciones profesionales adaptar y desarrollar cursos curriculares, materiales de capacitación in situ para estudiantes y profesionales de temas sobre la gestión o preservación de documentos de archivo digitales. Las universidades y asociaciones profesionales son libres de adaptar los materiales y de desarrollar, en su propio contexto, sus cursos específicos curriculares o paquetes de capacitación...

1.4 Objetivos

Los módulos tienen los siguientes objetivos:

- Proporcionar recursos educativos basados en investigación de vanguardia sobre temas de documentos de archivo digitales para beneficio de miembros de asociaciones profesionales relacionadas con esa temática;
- Proporcionar a archivistas y gestores de documentos de archivo el conocimiento teórico y procedimental así como habilidades estratégicas necesarias para desarrollar, implementar y supervisar un sistema de gestión o de preservación de documentos de archivo;
- Ilustrar conceptos teóricos con aplicaciones prácticas a través de ejemplos reales extraídos de casos de estudio, basados en contextos administrativos y tecnológicos específicos;
- Proporcionar contenidos y estructura a programas educativos universitarios para implementar cursos sobre administración o preservación de documentos de archivo.

1.5 Alcance

Los *Caminos de los Documentos de Archivo Digitales : Tópicos en Preservación Digital* consta de los siguientes módulos:

Módulo 1: Introducción – Un Marco de Referencia Para la Preservación Digital

Módulo 2: Desarrollo de Política y Procedimientos Para la Preservación Digital

Módulo 3: Cultura Organizacional y sus Efectos en la Administración de Documentos de Archivo

Módulo 4: Un Resumen de Metadatos

Módulo 5: De *Ad Hoc* a Regulado – Estrategias de Valoración para Lograr el Control de los Documentos de Archivo Digitales en Ambientes Web Distribuidos

Módulo 6: Gestión y preservación de Correo Electrónico

Módulo 7: Gestión y preservación de Documentos de Archivo en Ambientes Web

Módulo 8: Introducción al Cómputo en la Nube

Cada módulo contiene algunos o todos los siguientes componentes:

- **Resumen** del tema y alcance del módulo;
- **Objetivos de Aprendizaje** y nivel esperado de conocimiento al final del módulo;
- **Metodología** o procedimientos para la aplicación y desarrollo del módulo;
- **Plantillas (cuando apliquen)** para facilitar la implementación del módulo;
- **Ejemplos, Caso(s) de Estudio o Escenarios (cuando apliquen)** que proporcionan ejemplos de situaciones reales acerca de los temas del módulo;
- **Ejercicios** de los puntos clave del aprendizaje;
- **Preguntas de Revisión** que optimicen la comprensión y entendimiento del tema
- **Recursos Adicionales** para cada tópico que incluye **lecturas, estándares** y otras plantillas para referencia.

Resumen del Programa en Conjunto			
1. Un Marco de Referencia para la Preservación Digital			Fundamentos
2. Desarrollo de Políticas y Procedimientos para la Preservación Digital			
3. Cultura Organizacional	4. Resumen de Metadatos	5. Estrategias de Valoración	Generalidades
6. Correo Electrónico	7. Sitios Web	8. Cómputo en la Nube	Propósito específico
Base de Datos Internacional de Terminología			Fundamentos

1.6 Base de Datos Internacional de Terminología

La terminología utilizada en los módulos es la se usa comúnmente por las comunidades que practican la administración archivística. Para asegurar un entendimiento generalizado y reducir un potencial riesgo de confusión que pudiese surgir de prácticas regionales o jurisdiccionales, estos módulos están apoyados por una base de datos de administración de documentos alimentada internacionalmente. Está disponible en www.Web-denizen.com/. También ciertos términos específicos están incluidos en pequeños glosarios en cada módulo.

Módulo 7: Gestión y Preservación de los Documentos de archivo en Ambientes Web

2 Introducción

La naturaleza ubicua de la Web ha tenido como resultado la proliferación de sitios en este entorno, muchos de los cuales contienen información y documentos de archivo valiosos para las organizaciones. Los sitios Web se han movido más allá de actuar como repositorios de información hasta convertirse en sitios donde los documentos de archivo se producen, tienen lugar las transacciones (trámites) y se lleva a cabo la divulgación. Esto plantea cuestiones únicas sobre la gestión documental y la preservación para estos ambientes. Debido a la naturaleza de la Web asegurar la autenticidad de los documentos de archivo que se producen y residen en sitios Web puede ser difícil.

Muchas de las grandes organizaciones han sido fundamentales en el desarrollo de métodos de captura, herramientas y métodos de preservación.¹ Muchas de estas herramientas y técnicas son fácilmente adaptables a las necesidades de organizaciones o programas archivísticos pequeños o medianos. El Archivo Internet se ha desarrollado como una de las soluciones de código abierto para operaciones de recolección o cosecha que no requieren un costo monetario, pero sí un conocimiento tecnológico suficientemente extenso. Los Archivos Nacionales del Reino Unido han llevado a cabo una investigación sobre los mejores medios de almacenamiento, una guía simple para archivar sitios Web, así como investigación sobre los formatos óptimos para la producción de datos. Los Archivos Nacionales de Australia han realizado investigación sobre los requisitos de metadatos clave para manejar de forma efectiva todos los documentos de archivo digitales, incluyendo los documentos de archivo de actividades basadas en Web. Ellos también han investigado sobre soluciones para archivar (integrar) evidencia de documentos de archivo basados en Web en sitios con cambios frecuentes cuando se tienen buscadores poco frecuentes.

2.1 Propósitos y Objetivos

El objetivo de este módulo es introducir cuestiones clave involucradas en la gestión y preservación de sitios Web, incluyendo la identificación de documentos de archivo en ambientes Web. Esto ayudará a los usuarios a reconocer las necesidades de gestión y preservación de estos documentos de archivo, identificar las cuestiones involucradas en el manejo de documentos de archivo en un ambiente de Web e introducir estrategias para la preservación de sitios Web.

¹ Entre las más grandes organizaciones que actualmente preservan Sitios Web están la Biblioteca del Congreso de los Estados Unidos, el Archivo Internet, los Archivos Nacionales del Reino Unido y los Archivos Nacionales de Australia.

2.2 Resultados del aprendizaje

Al completar este módulo el estudiante será capaz de:

- Identificar documentos de archivo en ambientes Web;
- Entender cuestiones clave involucradas en la gestión y preservación de documentos de archivo en ambientes Web;
- Entender las diferentes estrategias para la preservación de sitios Web ;
- Saber dónde localizar información adicional y recursos que facilitarán cómo puede usted manejar y preservar sus documentos de archivo en ambientes Web.

2.3 Terminología

Esta sección identifica y define conceptos/constructos clave que son utilizados a través de este módulo.



Vea la Base de Datos Internacional de Terminología del Consejo Internacional de Archivos (ICA por sus siglas en inglés) en <http://www.Web-denizen.com> para más terminología relevante para este módulo.

Compatibilidad retrospectiva: Que es compatible con modelos o versiones anteriores del mismo producto. Se dice que una nueva versión de un programa tiene compatibilidad con versiones anteriores si el mismo puede utilizar archivos y datos producidos con versiones más antiguas del mismo programa. Se dice que una computadora es compatible si puede correr el mismo programa o aplicación de manera semejante que la versión anterior de la computadora.

Métodos de colección del cliente: La fuente desde la cual el sitio Web es acumulado para su preservación. Las acumulaciones del cliente son reunidas vía un navegador Web.

Transferencia directa: Adquisición de una copia de los datos directamente desde la fuente original. Requiere acceso al servidor principal. Involucra copiar datos desde el servidor y transferirlos a la institución responsable de su recopilación.²

Recolección (cosecha) remota: La técnica más común para archivar un sitio Web utiliza los buscadores de este entorno para automatizar el proceso de acumulación de páginas Web. Los rastreadores de Web normalmente “ven” las páginas Web de la misma manera que los usuarios con su navegador y, -por tanto-, proporcionan un método simple comparativo para recolectar contenidos Web de forma remota.

² Adrian Brown, *Archiving Websites* (London: Facet Publishing, 2006).

Métodos de recolección desde el servidor: La fuente desde la cual el sitio Web es recolectado para su preservación. Las acumulaciones en un servidor son reunidas a través del servidor Web.³

Rastreador de Web: Es un programa de computadora que busca en la World Wide Web de forma metódica y automática.

Sitio Web espejo: Un espejo es una copia exacta de un conjunto de datos. Esencialmente trabaja como una “copia impresa” de un sitio Web. El proceso para espejo de Web produce una copia del original, pero no captura los datos asociados.

3 Estrategia de Preservación de Sitios Web

No existe una solución única definitiva para ser aplicada a la preservación de sitios Web. Las estrategias dependerán de una variedad de factores que incluyen la presencia (o ausencia) de documentos de archivo en el sitio, la propiedad del contenido, las capacidades tecnológicas, costos y habilidades de almacenamiento. Por lo tanto, existen varios planes de acción que pueden ser concebidos para la preservación en el largo plazo de un sitio Web institucional. Los planes de acción abarcan desde soluciones relativamente baratas que simplemente preservan instantáneas de sitios Web en el tiempo hasta soluciones sumamente técnicas promovidas por ciertas empresas de bases de datos, altamente efectivas ya que abordan el dinamismo de sitios Web subyacentes.

Un número de herramientas están disponibles para archivar sitios Web. La herramienta que se seleccione dependerá de qué tanta información desea preservar la organización, las habilidades técnicas del personal y un meticuloso análisis de riesgo.

Existen muchas consideraciones para una organización que se embarca en un programa de preservación de sitios Web. Los factores incluyen: habilidades tecnológicas, administración de derechos; capacitación, descripción de recursos, documentación y acceso; selección de formatos; verificaciones de validación; planes de recuperación en casos de desastre; medios de almacenamiento, estándares y método de captura de sitios Web.

3.1 Capacidades Tecnológicas

Algunas estrategias requieren un conocimiento intensivo del ambiente tecnológico a fin de que pueda ser implementado, mientras otros requieren un cantidad mínima de conocimiento para implantar y tener éxito. Es relativamente simple tratar los sitios Web que integran documentos estáticos e incorporan muy poca o ninguna interoperabilidad. Sin embargo, aquellos sitios que incorporan altos niveles de interactividad y comprenden páginas generadas dinámicamente su tratamiento es complejo y han probado ser difíciles de archivar de manera efectiva.

³ Ibíd.

Es importante que las personas responsables de la preservación de materiales digitales cuenten con algún entendimiento de lo que todo esto involucra. El individuo responsable debe contar con suficientes conocimientos para tener un intercambio informado con quienes están involucrados en la estrategia de preservación, también debe ser capaz de establecer requisitos realistas a un tercero. Además, dentro de la organización, se deberán considerar las siguientes características cuando se evalúen las estrategias apropiadas de preservación. Entre ellas podemos distinguir:

- Tipo de sitio Web (por ej., estático o dinámico);
- Ubicaciones en espacio del servidor y disponibilidad;
- Capacidades de respaldo;
- Sistema de cómputo en uso.

3.2 Política/Requisitos de Gestión Archivística

Como profesional de documentos de archivo en su organización usted debe estar al tanto de los principios y prácticas de administración de documentos de archivo que se encuentran en práctica. Este módulo y los otros que lo acompañan son la oportunidad para que usted revise la efectividad de estas prácticas y la actualidad de cualquier documentación. Por ejemplo, usted puede desear revisar los requisitos de retención y disposición para ciertas series de documentos de archivo; actualizar políticas y procedimientos relacionados con las funciones de administración o reconsiderar las fortalezas y debilidades de las herramientas educacionales, las presentaciones o documentos que usted usa para la enseñanza a los empleados acerca de la importancia y prácticas de la administración de documentos de archivo.

Las políticas, procedimientos y criterios para un programa de preservación de sitios Web son críticos en el ambiente digital emergente. Éstos aseguran que los propósitos y objetivos de la institución son considerados y revisados cuidadosamente, que el desarrollo de colecciones sustenta la misión y prioridades institucionales y aseguran la rendición de cuentas de los organismos que financian y la comunidad en general. Los elementos a considerar en una política son: declaración de política, metas y objetivos de la política, documentos relacionados y/o legislación, alcance de la política, personas responsables de la implementación de la política, alcance de las colecciones, cobertura y una descripción de los tipos de recurso digital aceptados, criterios de rechazo, criterios de evaluación, viabilidad y niveles de la colección. Todos estos pueden ser divididos en más de una política.



Véase: “Los Caminos de los Documentos Digitales de Archivo: Módulo 2. Desarrollo de Políticas y Procedimientos para la Preservación Digital,” para abundar en el tema del desarrollo de una política.

Según la jerarquía de la política en su organización, la nueva documentación acerca del manejo de documentos de archivo en ambientes Web y preservación de sitios Web puede

ser independiente o puede formar parte de una política más extensa. Además de los elementos descritos en el Módulo de Política, la documentación sobre gestión y preservación de sitios Web deberá incluir elementos o secciones adicionales. En general, la estructura de su política estará basada en la información contextual de su organización.

3.2.1 Gestión Archivística

Todos los datos asociados con la preservación de sitios Web deberán estar incluidos en las tablas de retención que regulan los documentos de archivo de la institución. Las páginas Web deberán estar sujetas a la misma gestión archivística que se aplica a otros documentos de archivo electrónicos, ya que proporcionan evidencia de las actividades en línea de la organización. Además, si se cuenta con fechas de disposición en la gestión archivística, la organización se beneficiaría en términos de los costos asociados con el almacenamiento. Para asegurar la accesibilidad de datos en el largo plazo es esencial que el medio de almacenamiento sea refrescado a intervalos regulares. Si la organización almacena cada iteración del sitio Web indefinidamente entonces los costos asociados con los medios de refrescamiento se dispararán en el tiempo según el crecimiento de los datos acumulados.

3.3 Metadatos

Los metadatos son la clave para manejar de forma efectiva todos los documentos de archivo, incluyendo los documentos de archivo basados en actividades Web. Los *Lineamientos para Archivar Recursos Web de Australia*⁴ describen los requisitos de metadatos requeridos para diferentes escenarios.

Para documentos de archivo individuales en sitios Web y para documentos de archivo basados en actividades Web, esto significa el uso de metadatos para describir:

- Fecha y hora de producción y registro del documento de archivo en un sistema de gestión archivística;
- Contexto organizacional;
- Formato de datos original;
- El uso que se hace del documento de archivo en el tiempo, incluyendo su colocación en un sitio Web;
- Mandatos que regulan la producción, retención y disposición de documentos de archivo; y
- Gestión de la historia del documento de archivo después de su producción, lo que incluye mantenimiento, preservación y disposición.

⁴ National Archives of Australia, “*Archiving Web Resources: Guidelines for Keeping Records of Web-based Activity in the Commonwealth Government*,” (Marzo 2001). Disponible en: http://www.naa.gov.au/Images/archweb_guide_tcm16-47165.pdf

Para copias o instantáneas de todas las colecciones de recursos Web, los metadatos deben incluir:

- Fecha y hora de captura;
- Vínculos al Identificador Uniforme de Recursos (URI por sus siglas en inglés), incluyendo información acerca de la versión y fecha del vínculo al URI especificado;⁵
- Detalles técnicos acerca del diseño del sitio Web;
- Detalles acerca de los programas o aplicaciones utilizados para producir recursos Web;
- Detalles acerca de las solicitudes (incluyendo los motores de búsqueda) que complementan los recursos Web; y
- Detalles acerca de la aplicación requerida por el cliente para ver el recurso Web.⁶

Se recomienda que la auditoría de metadatos se lleve a cabo cuando se emprenda un plan o programa de preservación de sitios Web. Esto asegurará que los recursos capturados tienen anexados los metadatos suficientes para preservar de forma efectiva la exactitud, autenticidad, fiabilidad, accesibilidad y disposición de los recursos y permitir el acceso, además de que las actividades de preservación se lleven a cabo.



Véase: “Los Caminos de los Documentos de Archivo Digital: Módulo 4: Un resumen de Metadatos” para abundar en información acerca de metadatos.

3.4 Gestión de Derechos/Derechos de Propiedad Intelectual

Las cuestiones relacionadas con los derechos de propiedad intelectual, tales como los derechos de copia (copyrights) y los derechos morales tienen un importante impacto en cualquier proceso de preservación digital. *“Los temas de derechos de propiedad intelectual en materiales digitales son... más complejos y significativos que aquellos en medios tradicionales y si no son atendidos pueden obstruir e inclusive impedir las actividades de preservación.”*⁷ No se trata tan sólo del contenido, sino de los programas o aplicaciones asociados, los cuales pueden estar sujetos a derechos de propiedad intelectual. *“Simplemente el copiado (refrescado) de materiales digitales a otro medio, el*

⁵ Los lineamientos *“Australian Guidelines for Archiving Web Resources”* hacen distinciones entre un URI, un URL y un URN de este modo: el *Identificador Universal de Recursos (URI)* es simplemente un mecanismo de vocabulario indexado (namespace) de propósito general; el *Localizador Universal de Recursos (URL)* es una instancia del URI que es la dirección de algún recurso accesible mediante un protocolo como el HTTP; *El Nombre Universal de Recurso (URN)* es una instancia de URI, que a diferencia de un URL que es frágil, está garantizado a permanecer disponible (Jon Udell, *Practical Internet Groupware* (Sebastapol, CA: O’Reilly, 1999), p. 471.)

⁶ *“Archiving Web Resources: Guidelines for Keeping Records of Web-based Activity in the Commonwealth Government”* de los National Archives of Australia, p. 17-18.

⁷ Maggie Jones y Neil Beagrie, *“Preservation Management of Digital Materials. A Handbook”* (London, UK: The British Library, 2001), 32.

encapsulamiento del contenido y los programas para emulación o migración del contenido a un equipo o programas nuevos involucra actividades que pueden infringir los derechos de propiedad intelectual a menos de que existan exenciones regulatorias o se hayan obtenido permisos específicos de los propietarios de los derechos.”⁸ Debido a la naturaleza de los materiales digitales, las estrategias para la preservación y el acceso continuados van a necesitar la migración de materiales a nuevas formas o a una emulación del ambiente operativo original. Tales actividades requieren permisos de los propietarios de los derechos para emprender tales estrategias.

Un área que podría convertirse potencialmente problemática es el área de las Leyes de Derechos de Autor. Según la Web Canadiense de Información Patrimonial (CHIN, siglas en inglés) “*El derecho de autor protege la expresión de ideas que están fijas en cualquier forma de medio.*”⁹ Esto incluye varios componentes del sitio Web, tales como las imágenes que aparecen en cierto sitio así como el código de programación de cómputo subyacente.

El derecho de autor protege la mayoría de las creaciones tales como trabajos literarios, teatrales y artísticos, grabaciones de sonido y trabajos audiovisuales. Las fotografías son consideradas trabajos artísticos. Los programas de computadora en forma de código subyacente en una página Web han sido identificados como trabajos literarios y por tanto están protegidos por el derecho de autor. Con excepción de trabajos producidos en el curso de las obligaciones de un empleado o donde el derecho de autor ha sido asignado por escrito a alguien más, el autor de trabajo es el propietario del derecho de autor.¹⁰

Los propietarios de derechos de autor deben ser establecidos y los permisos garantizados antes de iniciar un programa de preservación Web.

3.5 Desarrollo y Capacitación del Personal

La capacitación cuidadosamente diseñada para el personal y el desarrollo profesional continuo puede jugar un papel clave en el manejo exitoso de cualquier programa de preservación. Todos aquellos responsables de la preservación digital deben contar con cierto grado de conocimiento sobre el tema. El desarrollo y capacitación del personal puede ir desde mantenerse actualizado con la literatura y los nuevos desarrollos hasta participar en talleres o módulos de capacitación presentados por varias instituciones tales como asociaciones archivísticas e instituciones educativas.¹¹

⁸ *Ibíd.*

⁹ Rina Elster Pantalony, “*Protecting your Interests: a legal guide to negotiating Website development and virtual Exhibition Agreements*”. (Ottawa, Canada: Minister of Public Works and Governments Services Canada, 1999), 13.

¹⁰ *Ibíd.*

¹¹ “*The Society of American Archivists*” es una institución que organiza numerosos talleres y seminarios vía la Web. Para una vista de su programa de actividades actuales véase: <http://saa.archivists.org/Scripts/4Disapi.dll/4DCGI/events/ConferenceList.html?Action=GetEvents>.

3.6 Descripción del Recurso, Documentación y Acceso.

Alguna forma de descripción de clasificación es esencial a fin de manejar cualquier colección archivística y hacerla accesible a usuarios; esto no es diferente en una colección digital. Los grandes estándares para catalogación, tales como MARC 21 e ISAD(G), han sido aplicados exitosamente para la descripción de sitios Web que se archivan. La catalogación y clasificación de materiales archivados permite a los usuarios acceder a los mismos.

Los recursos deben ser suministrados con documentación apropiada y suficiente para satisfacer los requisitos para uso informado por parte de los miembros de la comunidad de investigación. La documentación debe estar relacionada tanto con el contenido como con el formato técnico del recurso. La documentación debe proporcionar también información acerca del contexto en el cual los recursos fueron producidos y mantenidos antes de su preservación y acerca de las relaciones del recurso digital y otras fuentes de información.

3.7 Plan de Recuperación en Caso de Desastre

El desarrollo de un plan de recuperación en caso de desastre que esté basado en principios sólidos debe contar con la aprobación de la dirección y debe poder ser activado por el personal capacitado. Esto reducirá ampliamente la severidad del impacto hacia la organización en caso de desastre. El plan necesitará orientarse a la restauración tanto del contenido del archivo como de la infraestructura técnica y operacional requerida para soportarlo. Los elementos que deberá incluir el plan son:

- Asegurar que el personal sea capacitado en procedimientos contra desastres;
- Producir copias de archivos de recursos de datos cada vez que una colección de materiales se lleva a cabo; almacenar copias archivadas en múltiples medios.
- Almacenar copias archivadas dentro y fuera de sitio;
- Contar con documentación completa de la infraestructura de equipo y programas de cómputo así como los procedimientos operativos y manuales;
- Contar con copias de todos los programas de cómputo requeridos para operar los sistemas.

También es importante probar el plan para identificar cualquier cuestión que pueda haber sido omitida antes de que un evento de desastre ocurra. Esto también ayuda al personal a familiarizarse con esos procedimientos con anterioridad. Como con la mayoría de las políticas, se recomienda que el plan de recuperación en caso de desastre sea revisado según haya cambios en sistemas y circunstancias.

Ejercicios:

- ¿Tiene su organización actualmente un plan de recuperación en caso de desastre?
- Analice cómo los recursos de su organización podrían caber en este plan.

3.8 Verificaciones de validación

Una vez que el sitio Web ha sido integrado y transferido al entorno de la institución de archivo histórico, se deben llevar a cabo verificaciones para asegurar que todas las partes del sitio Web integrado están trabajando como deben. Las verificaciones incluyen, -entre otras: teclear manualmente todos los hipervínculos; teclear al azar sobre los hipervínculos o emplear el uso de una aplicación de prueba de estos vínculos para ayudar a automatizar el proceso de verificación en el sitio Web que se está trabajando,¹² verificando que los archivos pueden ser leídos, están completos y exactos así como la existencia de funcionalidad dentro de los archivos. Estas verificaciones deben ser llevadas a cabo cuando un sitio Web se ha archivado a fin asegurar que el contenido y las estructuras de los recursos de datos están intactos.

3.9 Formatos de Archivo

En cualquier programa de preservación de sitios Web (como con cualquier programa de preservación digital) se recomienda que los formatos de archivo aceptados sean definidos antes de emprender cualquier estrategia de acumulación. La adopción de un sólo formato asegura que los costos de sustentabilidad se vean reducidos cuando el formato de archivo seleccionado es incorporado desde el proceso de producción de los documentos de archivo.

“Se ha convertido en una práctica común en los repositorios de documentos de archivo digital, incluyendo archivos históricos, el hecho de aceptar ciertos formatos de archivos digitales para la preservación en el largo plazo y rechazar otros”¹³. Las encuestas de instituciones respecto de las especificaciones de formatos de archivo muestran que existe una plétora de definiciones, formatos aceptables o inaceptables, e iniciativas de preservación para formatos de archivos.¹⁴ El “Diccionario para la Preservación de

¹² Véanse como ejemplos: Link Checker Pro: <http://www.link-checker-pro.com/>; Site Audit: http://www.blossom.com/site_audit.html; Cyber Spyder Link Test: <http://www.cyberspyder.com/cslnkts1.html>; Link Sleuth: <http://home.snafu.de/tilman/xenulink.html>.

¹³ Evelyn Peters McLellan, “General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation,” InterPARES 2 Project (March 2007), 1. Disponible en: http://www.interpares.org/display_file.cfm?doc=ip2_gs11_final_report_english.pdf.

¹⁴ *Ibíd.*

Metadatos PREMIS” brinda la definición más útil: “una estructura específica preestablecida para la organización de un archivo digital o de una cadena de bits. Esta estructura preestablecida incluye la forma en que están codificados los datos y la forma en la cual los bits son interpretados para producir texto, imágenes y sonido.”¹⁵ Esto es importante, ya que debe hacerse énfasis en el hecho de que es esencial especificar los formatos de archivo aceptables para un repositorio específico. “Algunos tipos de codificación son sinónimo de formatos específicos de archivo; por ejemplo, la codificación mp3 es utilizada para codificar un formato de archivo .mp3”¹⁶ Esto parece suficientemente simple de entender, pero se hace más complicado. Si tomamos por ejemplo los archivos de texto plano “estos pueden tener formatos con codificaciones diferentes, ya que pueden ser codificados como ASCII, EBCDIC o Unicode, entre un gran número posible de variantes.”¹⁷ Si sólo el texto plano tiene tres tipos diferentes de codificación, es obvio que los archivos de imagen y música serán mucho más complicados a este respecto. “La codificación puede ser problemática en un formato de archivo de audio y video debido a que la codificación óptima para almacenamiento y transmisión con frecuencia conlleva la compresión (remoción de bits de los archivos digitales para reducir su tamaño), lo cual con frecuencia entorpece los esfuerzos de preservación.”¹⁸ Dificultades adicionales vienen a debate de formatos de archivos: “la cuestión de codificación es más complicada por el hecho de que TIFF, WAV y AVI así como otros formatos de imagen y audiovisuales comunes no son ‘formatos’ de archivo propiamente dichos, sino que son ‘formatos envolventes’ (también llamados formatos contenedores), los cuales están diseñados para combinar cadenas de bits de naturaleza distinta en un sólo archivo de computadora.”¹⁹ La codificación, las combinaciones de compresión y cadenas de bits: todas complican la forma en que los formatos de archivo son preservados a largo plazo. Éstas son también razones de por qué muchas instituciones piden formatos abiertos que estén bien documentados para asegurar que suficiente documentación esté disponible para brindar a la institución colectora una oportunidad para preservar los documentos de archivo a largo plazo.

Adrian Brown de los Archivos Nacionales del Reino Unido ha identificado criterios a considerar cuando se seleccionan los formatos de producción de datos; entre éstos están:

- Ubicuidad
- Asesoría o soporte
- Confidencialidad
- Calidad de la documentación
- Estabilidad
- Fácil identificación
- Derechos de propiedad intelectual
- Soporte de metadatos

¹⁵ *Ibíd*, 2.

¹⁶ *Ibíd*.

¹⁷ *Ibíd*.

¹⁸ *Ibíd*.

¹⁹ *Ibíd*.

- Complejidad
- Interoperabilidad
- Viabilidad
- Reusabilidad

Aunque la investigación no recomienda en especial tipos actuales de archivos, es importante tener en mente estos criterios cuando se seleccionan los formatos de archivo.²⁰

Es importante que las organizaciones productoras y preservadoras desarrollen una política que establezca los tipos de formatos de archivo que son aceptables para preservar. Mediante la restricción del rango de formatos de archivo que una institución acepta recibir y manejar, la organización puede estar segura que los formatos de archivo que acumula se adhieren a los criterios establecidos arriba y que se adhieren a los estándares actuales. Si son acumulados “buenos” formatos de archivo, las dificultades en su preservación se reducirán, así como los costos.

*“Muchos formatos son propietarios, esto es, son propiedad de un dueño quien por razones comerciales no está dispuesto a proporcionar acceso a la documentación acerca de éstos y puede requerir que se le pague una tarifa por su uso.”*²¹ Ésta es una razón por la cual la mayoría de los expertos recomiendan formatos de archivo que se adhieran a estándares abiertos. Ésta también es una razón de por qué muchos registros de formatos de archivo se han desarrollado. Los registros existen para proporcionar información fiable y detallada acerca de los formatos de archivo. Ejemplos de registros de formatos de archivo incluyen a: PRONOM²² y el Registro Global de Formatos Digitales (*Global Digital Format Registry*).²³ En abril de 2009 la iniciativa del Registro Global de Formatos unió esfuerzos con la iniciativa PRONOM de los Archivos Nacionales del Reino Unido bajo un nuevo nombre, el “*Registro Unificado de Formatos Digitales*” (UDFR por sus siglas en inglés). El UDFR mantendrá los requisitos y casos de uso compilados por el Registro Global de Formatos Digitales y será integrado con la base de datos de aplicaciones y formatos de PRONOM.²⁴

La organización responsable de acumular documentos de archivo puede ayudar a promover la producción sólida de documentos de archivo mediante la publicación de aquellos formatos de archivo que tienen mayores posibilidades de ser sustentables en un período y alentando a la producción de documentos de archivo que utilice estos formatos. Otra alternativa para la institución que acumula es convertir todos los materiales digitales

²⁰ Adrian Brown, “*Selecting File Formats.*” Disponible en:

<http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>

²¹ Ross Harvey, “*Preserving Digital Materials*” (Munich: K. G. Saur, 2005), 141.

²² PRONOM es un registro de formatos de archivo establecido por los Archivos Nacionales del Reino Unido para proporcionar y manejar información acerca de los formatos y aplicaciones de cómputo utilizados. El sitio Web de PRONOM puede encontrarse en: www.nationalarchives.gov.uk/pronom.

²³ El Registro Global de Formatos Digitales fue también desarrollado para apoyar a la preservación digital. Disponible en: <http://www.gdfr.info/>

²⁴ El Registro Unificado de Formatos está disponible en: <http://www.udfr.org/>

preservados al formato de archivo seleccionado una vez que el material está en el archivo histórico.

3.10 Medios de Almacenamiento²⁵

Sea cual sea el método de captura, el sitio Web necesita ser preservado y almacenado en un medio o soporte electrónico digital relativamente estable. En la actualidad, ningún medio electrónico digital puede ser considerado como archivístico debido a que los períodos en los que tales medios han sido probados son todavía cortos, además de preocupaciones respecto de la obsolescencia tecnológica que son resultado de cambios rápidos en este entorno. Los dispositivos para almacenamiento están en desarrollo y cambio continuo. El actual “estado del arte” de este ambiente puede ser obsoleto en un período de tan sólo cinco años y es simplemente imposible mantenerlo por veinte años. Los medios electrónicos no son tan permanentes como se piensa con frecuencia. Los fabricantes aseguran satisfactoriamente tiempos de vida largos para sus medios,²⁶ pero la experiencia práctica sugiere que una figura realista para la vida de una cinta magnética puede ser de quince años y para un disco compacto veinte años; todo esto según la calidad del original, almacenamiento, manejo y uso. Y aun si el tiempo de vida es mayor, el equipo para leerlo puede no estar disponible. Para muchos medios, una imperfección que aparece después de un tiempo puede hacer que todo el medio sea inutilizable.²⁷ Por tanto, cualquiera que sea el medio seleccionado para almacenamiento necesitará verificarse periódicamente y/o refrescarse para contrarrestar la pérdida de datos.²⁸

Una variedad de factores afectan la longevidad de los medios electrónicos, incluyendo las condiciones de almacenamiento, la calidad de los productos utilizados y la composición de los productos debido a la disponibilidad de mejores materiales con el tiempo. Por tanto, es muy difícil predecir la longevidad. El Instituto Canadiense de Conservación ha armado una tabla que proporciona estimaciones de longevidad para varios medios de almacenamiento.

²⁵ En este reporte presentamos el medio o soporte básico de almacenamiento para guardar documentos electrónicos. Es posible crear un repositorio para materiales digitales. Si usted requiere más información revise la norma ISO: 14721:2003, comúnmente conocida como el estándar ISO:14721, más comúnmente conocido como el “Modelo de Sistemas de Información de Archivo Abierto” (OAIS, por sus siglas en inglés) así como OCLC y NARA. “*Trustworthy Repositories Audit & Certification: Criteria and Checklist*” Version 1.0, 2007. Disponible en: <http://www.crl.edu/PDF/trac.pdf>

²⁶ En 1995 una investigación de Kodak en sus CD’s reescribibles reportó un período de vida útil de 217 años bajo condiciones específicas. Disponible en: <http://www.cd-info.com/archiving/kodak/index.html>

²⁷ Jim Liden Sean Martin, Richard Masters y Roderic Parker, “*The large-scale archival storage of digital Objects*,” DPC Technology Watch Series Report, 04-03, Febrero 2005.

²⁸ Véase The National Archives of the UK’s Digital Preservation Guidance Note: 2, “*Selecting Storage Media for Digital Preservation*,” by Adrian Brown, Head of Digital Preservation Research, Agosto 2008 Disponible en: <http://www.nationalarchives.gov.uk/documents/selecting-storage-media.pdf>

Longevidad pronosticada de medios electrónicos²⁹

Tipo de medio	Longevidad pronosticada
Discos magnéticos	
Discos duros	2–5 años
Disquetes flexibles	5–15 años
Cintas magnéticas	
Digitales	5–10 años
Analógicas	10–30 años
Discos ópticos	
CD-RW, DVD-RW, DVD+RW	5–10 años
CD-R (cianina y colorantes azoicos)	5–10 años
Audio CD, DVD, cine	10–50 años
CD-R (colorante de ftalocianina, capa metálica de plata)	10–50 años
DVD-R, DVD+R	10–50 años
CD-R (colorante de ftalocianina, capa metálica de oro)	>100 años
Otros discos ópticos	
MO, WORM, etc.	¿10–25 años?
Dispositivos tipo flash	desconocido

Se recomienda por lo tanto que el sitio Web archivado sea almacenado en varios soportes, -por ejemplo en un disco duro y en un DVD-R-, y guardado en los archivos para contrarrestar estos problemas de almacenamiento y ayudar a asegurar el acceso de los datos almacenados en el largo plazo.

En la determinación del tipo de medio de almacenamiento para almacenar materiales digitales es necesario considerar cierto número de factores. Estos factores incluyen, longevidad, capacidad, viabilidad, costos y sustentabilidad, según lo documenta Adrian

²⁹ Canadian Conservation Institute, “*Electronic Media Collections Care for Small Museums and Archives*”. Disponible en: http://www.cci-icc.gc.ca/headlines/elecmediacare/index_e.aspx

Brown de los Archivos Nacionales del Reino Unido.³⁰ Brown expone una tabla de puntuación que compara los tipos de soportes o medios comunes:

Medio	CD-R	DVD-R	Disco Duro	Memorias y tarjetas flash (Dispositivos USB)	Cinta de Almacenamiento de datos (Linear Tape Open [LTO])
Longevidad	3	3	2	1	3
Capacidad	1	3	3	2	3
Viabilidad	2	2	2	1	3
Obsolescencia	1	2	2	2	2
Costo	3	3	1	3	3
Susceptibilidad	1	1	3	1	3
Total	11	14	13	10	17

Según esta tabla, las soluciones con puntajes más altos son la cinta de almacenamiento LTO y el DVD-R, junto con la opción de disco duro como un tercero cercano; Brown aconseja:

“En situaciones donde las copias múltiples de datos son almacenados en medios separados, puede ser ventajoso usar diferentes tipos de medios para cada copia, utilizando preferentemente diferentes tecnologías base (por ejemplo magnética y óptica). Esto reduce en general la dependencia a la tecnología de los datos almacenados. Donde se usa el mismo tipo de medio para copias múltiples, se deberán usar diferentes marcas o lotes en cada caso a fin de minimizar los riesgos de pérdida de datos debido a problemas con fabricantes o lotes específicos”.

Joe Iraci, del Instituto Canadiense de Conservación tiene comentarios adicionales respecto las diferencias del medio de almacenamiento. En cuanto al uso de medios ópticos para almacenamiento, Iraci señala: *“el tipo de disco utilizado y la forma como es grabado impacta enormemente en la longevidad”.* Él destaca que *“las cintas digitales tienen corta duración y necesitan ser migradas/refrescadas cada 5-10 años”* y advierte que *“los discos duros no son para el almacenamiento en el largo plazo ya que los datos necesitan moverse a un nuevo disco duro cada 2 a 5 años”* y nos recuerda *“seguir con tecnologías ya generalizadas y evitar nuevas tecnologías tales como discos Blue-Ray, almacenamiento holográfico [y] memorias flash”.* Iraci también señala que *“con todos*

³⁰ The National Archives, “Digital Preservation Guidance Note 2: Selecting Storage Media for Long-Term Preservation,” Agosto 2008. Disponible en: <http://www.nationalarchives.gov.uk/documents/selecting-storage-media.pdf>

los medios digitales, los respaldos son críticos a fin de evitar una pérdida repentina de información.”³¹

Las investigaciones como las realizadas por Adrian Brown y el Instituto de Conservación Canadiense son invaluableles cuando hay que decidir el medio para el almacenamiento institucional de documentos de archivo. Es claro que se debería elegir una variedad de medios y que aún con el almacenamiento y manejo correctos el medio debe ser verificado y refrescado regularmente.

3.11 Normas

Cierto número de estándares están relacionados con el archivado de sitios Web. HTML y XML son el núcleo de tecnologías reconocidas como estándares en la forma de las recomendaciones del W3C³². Existen dos estándares en el área de gestión de documentos de archivos, ISO 15489-1/2:2001, el cual establece los estándares para la práctica de la gestión archivística e ISO 23081-1:2006, quien establece estándares para la gestión de metadatos para documentos de archivo.

ISO 14721:2003 establece la norma que define los requisitos fundamentales para un sistema de preservación digital. Más conocido como el “*Modelo de Referencia para un Sistema Abierto de Información Archivística*” (OAIS por sus siglas en inglés), sus conceptos y metodología han sido ampliamente adoptados por una audiencia internacional. Forma la base para un esquema de certificación de repositorios digitales de confianza.



Véase “Los Caminos de los Documentos de Archivo Digitales: Módulo 1: Introducción - Un Esquema para la Preservación Digital” para abundar en el modelo de referencia OAIS.

La ISO 19005-1: 2005 o el estándar PDF/A ha abordado la necesidad de formatos de archivo digitales abiertos. Este estándar es “*un formato de archivo basado en PDF, conocido como PDF/A, el cual ofrece un mecanismo para representar documentos electrónicos de tal forma que preserva su apariencia visual al paso del tiempo, con independencia a las herramientas y sistemas utilizadas para producir, almacenar o presentar los archivos.*”³³

³¹ Correo electrónico de Joe Iraci a Randy Preston, Mayo 20, 2009.

³² W3C, o World Wide Web Consortium o el Consorcio de la Web Mundial es un consorcio internacional donde las organizaciones miembros, personal de tiempo completo y el público trabajan juntos para desarrollar estándares para la Web.

³³ ISO-19005-1 – Administración de Documentos – Formato de archivo para la preservación en el largo plazo de un documento electrónico - Parte 1: Uso de PDF 1.4 (PDF/A-1).

3.12 Mantener los Documentos de Archivo basados en la Web a través del Tiempo

Asegurar la accesibilidad de los materiales basados en la Web al paso del tiempo plantea las mismas cuestiones de accesibilidad que rodean otros documentos de archivo electrónicos. Existen pasos que pueden ser tomados para mitigar estas cuestiones, incluso el asegurar que los materiales son manejados cuidadosamente (inclusive mantener la confianza de los documentos de archivo en la Web, e identificar y mitigar la gestión de riesgos), tales como planes para la obsolescencia, uso de estándares ampliamente sustentados, implementación de medidas de seguridad para proteger en contra de alteraciones deliberadas o accidentales y asegurar el control ambiental y supervisión. La mayoría de estos pasos han sido discutidos en partes previas de este documento, pero es prudente insistir de nuevo sobre su importancia.

Administración Cuidadosa: Esto puede incluir: mantener matrices “masters” y almacenarlos en una ubicación separada; implementar el uso de XHTML y evitar el uso de etiquetas no estandarizadas de HTML; refrescar el medio de almacenamiento periódicamente; llevar a cabo verificaciones al azar para asegurar la accesibilidad a los datos.

Planeación para la Obsolescencia: Planear la obsolescencia asegurando que los documentos de archivo pueden ser copiados, refrescados o migrados. Cualquier actividad de preservación como las mencionadas deberá estar bien documentada en los metadatos de gestión archivística, incluyendo cualquier pérdida de funcionalidad, contenido o apariencia.

Uso de Normas y estándares: Se destaca la importancia del uso de normas y/o estándares.

Medidas de Seguridad como Medio para Proteger Datos: Es importante construir medidas de seguridad en la preservación de la Web que protejan los datos de alteraciones deliberadas o accidentales. Las medidas pueden ser tan simples como mantener los datos preservados en un ambiente seguro con acceso controlado para permitir sólo a personas autorizadas acceder a los datos y proporcionar acceso sólo para leer los datos preservados, no así para modificarlos.

Control Ambiental y Monitoreo: La mejor práctica establece que los medios almacenados deben ser mantenidos a niveles óptimos de temperatura y humedad; los medios deben estar protegidos contra campos magnéticos; se debe contar con unidades de filtración de aire para proteger contra impurezas del mismo; prohibir el consumo de alimentos en el área de almacenamiento; y planeación en caso de desastre.



Véase la “Guía NARA sobre Manejo de Documentos de Archivo en la Web”, disponible en: <http://www.archives.gov/records-mgmt/policy/managing-Web-records-index.html>

3.13 Métodos de Captura de Sitios Web /Herramientas

Existen dos tipos de sitios Web: estáticos y dinámicos. Un sitio Web estático está compuesto de una serie de páginas Web construidas previamente, mismas que están vinculadas desde al menos otra página. Un sitio Web dinámico genera páginas Web sobre la marcha a partir de pequeños elementos de un contenido que puede estar guardado en una base de datos, recogido desde fuentes externas e insertado en la página Web, o generado mediante programas que responden de forma diferente según factores tales como la fecha u hora en la que se accedió a la página Web.

Actualmente existen tres opciones disponibles para capturar sitios Web. Los métodos para captura dependen de qué tanta información desea preservar la institución que lo conservará. Esta información incluye funcionalidad, metadatos y el grado de autenticidad, fiabilidad y exactitud que la institución desea preservar. Las tres opciones son: transferencia directa, recolección remota y sitio Web espejo.

3.13.1 Transferencia Directa

La única forma para recrear completamente un sitio Web en un ambiente de preservación es a través de la transferencia directa de datos. La transferencia directa opera mediante la adquisición de una copia de los datos directamente desde la fuente original. Esto requiere acceso directo al servidor primigenio de alojamiento de la Web. Así, la transferencia directa consiste en copiar los archivos seleccionados desde el servidor y transferirlos a la institución que los colecciona. Para garantizar la funcionalidad continuada será necesario llevar a cabo ajustes menores al sitio preservado.³⁴ Para asegurar que el sitio Web preservado es lo más auténtico posible, será necesario implementar una recreación del ambiente técnico en el cual reside el sitio Web en el entorno archivístico. Esto significa que la base de datos o el sistema de administración de contenidos requerirán ser instalado en el ambiente archivístico, junto del servidor de Web y los programas del motor de búsqueda necesarios. La transferencia directa es el único método que toma en consideración la naturaleza dinámica de un sitio Web y es la única forma para preservar formas posibles de datos generados dinámicamente. Sin embargo, la implementación y soporte de tal método requerirá de la disponibilidad de personal con habilidades técnicas para instalar y mantener el sistema.

³⁴ Por ejemplo, los hipervínculos dentro de un sitio archivado pueden necesitar ser ajustados desde vínculos absolutos a vínculos relativos y el motor de búsqueda apropiado (el utilizado en el ambiente original) debe ser instalado en un nuevo ambiente para asegurar que también la funcionalidad de búsqueda es preservada. Para una explicación más amplia véase: Adrian Brown, “*Archiving Websites*”.

3.13.2 *Recolección Remota*

La solución de recolección remota ofrece tres alternativas: rastreo directo automatizado del sitio Web, rastreo de “instantáneas” con bitácoras adicionales mantenidas por el archivista para respaldar los datos extraídos en la instantánea, y la tercerización del proceso. Los métodos de colección por recolección remota como alternativa deben ser entendidos con la advertencia de que tales métodos de recolección no capturan la totalidad de todas las posibilidades de página de la Web, mismas que pudieran generarse por solicitud de un usuario cuando el sitio Web identificado para su captura es un sitio dinámico con una base de datos subyacente (back end) utilizada para albergar la información generada sobre la marcha. También, el uso de este método puede tener como resultado la presencia de hipervínculos rotos dentro del ambiente de datos copiados, ya que las páginas pueden contener vínculos a contenido que necesita ser generado sobre la marcha para presentarse al usuario. Otra pérdida de datos que podría ocurrir puede ser la pérdida de gráficas y el diseño de plantillas.

Una instantánea de un sitio Web usualmente involucra la creación de una copia completa y exacta del sitio Web de una organización en un punto específico del tiempo. Una instantánea deberá incluir todos los aspectos del sitio Web para asegurar que un sitio funcional completo puede ser recreado. Una instantánea deberá incluir guiones, programas, complementos (plug ins), y las aplicaciones de búsqueda, componentes que en conjunto hacen que la instantánea sea completamente funcional.

Un estándar para rastreo de Web debe ser conducido utilizando un programa de recolección de sitio Web de código abierto, como las herramientas gratuitas GNU Wget o Heritrix.

Las ventajas de un rastreador de código abierto para archivo de sitio Web son las de no ser propietario y por lo tanto no se incurre en erogaciones financieras. Un rastreador automatizado de Web puede coleccionar datos tan frecuentemente como lo desee la institución; inicialmente el rastreador podría establecerse para rastrear todo el sitio, posteriormente los rastreos subsecuentes pueden coleccionar datos sólo de páginas que hayan sido actualizadas a partir del rastreo previo.

Para preservar instantáneas de un sitio Web en cierto momento del tiempo, la institución necesita rastrear el sitio Web sólo una o dos veces al año. Esta frecuencia sin embargo, obviamente no captura cada cambio hecho al sitio Web y puede perder algunas de las actividades documentadas que se presenten. El rastreador de Web puede ser implementado para realizar rastreos poco frecuentes del sitio Web. Así, “copias instantáneas” del sitio Web como un todo son tomadas (asegurando que la funcionalidad de los vínculos internos no sea destruida y se mantienen) y mientras tanto, para asegurar que la evidencia necesaria es capturada se mantiene una bitácora de cambios que determina cuándo y cómo los documentos o las páginas Web son removidos, reemplazados o actualizados. Si los propósitos de rendición de cuentas y mantenimiento son importantes a fin de que los documentos de archivo del contenido del Sitio Web y sus

cambios sean realizados y mantenidos, entonces esta es una opción viable y barata.³⁵ Una vez más, los metadatos son la clave para manejar todos los documentos de archivo de forma efectiva, incluyendo los documentos de archivo basados en la actividad Web. (Véase el Módulo previo sobre metadatos).

Existe la opción de tercerizar o subcontratar la captura y almacenamiento de sitios Web a compañías con base en un pago. Los servicios tales como WAS, Web Archiving System desarrollado por la Biblioteca Digital de California y el Proyecto Archive-It operado por el *Internet Archive* proporcionan servicios de captura y almacenamiento de Web a organizaciones que desean preservar sus sitios Web. Es importante mencionar que con esta opción los datos almacenados por compañías que hospedan estos servicios pueden estar distribuidos a través del planeta y sujetos a una variedad de leyes y reglamentos jurisdiccionales. Es importante que antes de emprender un proyecto de este tipo se esté perfectamente al tanto de su esquema legislativo y regulatorio respecto de protección de datos y acceso a la información.

3.13.3 Sitio Web Espejo

El sitio Web espejo es una opción que copia a imagen el sitio Web, pero no captura los metadatos asociados necesarios para preservar de forma efectiva el contenido digital del sitio Web. Un espejo es una copia exacta de un conjunto de datos. Esencialmente opera como una “copia impresa” pero en forma digital del sitio Web. El espejo de sitios se lleva a cabo por varias razones, una de éstas para preservar un sitio o una página Web.

Como ya se señaló arriba, al no capturar los metadatos asociados con cada archivo de la página Web es una buena opción cuando el archivo histórico desea sólo preservar evidencia del sitio Web en cuestión. Esta solución debe ser entendida con la previsión de que no hay metadatos de evidencia de los documentos de archivo actuales que puedan aparecer en el sitio. Por lo tanto, no es recomendado si el archivo histórico que los colecciona desea preservar evidencia de los documentos de archivo que aparecen en el sitio Web.

3.13.4 Herramientas de Captura de Web

El rastreador de código abierto HTTrack ha sido utilizado de forma efectiva en otras instituciones archivísticas.³⁶ El HTTrack es gratuito y es una herramienta para navegar fuera de línea y de uso fácil. Permite al usuario bajar un sitio Web desde Internet a un directorio local, construir de forma recurrente todos los directorios, copiar HTML, imágenes y otros archivos desde el servidor al directorio local. HTTrack arregla la estructura de enlaces relativos del sitio original. Permite a los usuarios simplemente abrir

³⁵ El rastreo de Web con opción de registro fue investigado utilizando la obra “*Archiving Web Resources: Guidelines for Keeping Records of Web-based Activity in the Commonwealth Government*” de los Archivos Nacionales de Australia. Es un documento de gestión documental de gobierno publicado en Marzo de 2001 y puede verse en: http://www.naa.gov.au/Images/archWeb_guide_tcm2-903.pdf

³⁶ Correo electrónico enviado a la lista *Management & Preservation of Electronic Records Listserv*, Abril 3, 2009, por parte del Archivista de Documentos de Archivo Electrónicos del Departamento de Archivos y Bibliotecas de Kentucky.

una página con el sitio Web “reflejado” en su rastreador y buscar el sitio de vínculo a vínculo, como si se viera en línea.³⁷ Los archivistas que buscan preservar contenidos de Web en ambientes de Microsoft/Windows han usado este recolector con éxito.



Véase “*Practical E-Records*” para una revisión de las herramientas *HTTrack*, *GNU Wget*, *Heritrix* y *Web Archiving Service*. Disponible en: <http://e-records.chrisprom.com/?tag=Website-harvesting>. Todo esto fue revisado con base en los siguientes criterios: *Instalación / Configuración / Plataformas Soportadas / Funcionalidad / Fiabilidad; Usabilidad; Escalabilidad, Documentación; Interoperabilidad / Metadatos de soporte; Flexibilidad / Personalización; Licencia / Soporte / Sustentabilidad / Comunidad.*

Además, la *Herramienta de Captura de Web* de Adobe (*Web Capture*) convierte las páginas Web a archivos de PDF para crear versiones PDF de la página Web. Es simple para usar y por tanto fácil de enseñar al personal. Es posible capturar todo un sitio Web usando esta herramienta. No sólo todos los vínculos continúan operando en el PDF, también pueden vincular el contenido local dentro del PDF cuando es pertinente, de tal forma que se puede navegar en el sitio fuera de línea. *Web Capture* puede ser invocada a través de la barra de herramientas de Acrobat en el Internet Explorer en Windows y a través de la aplicación de Adobe Acrobat 9 en plataformas Windows y Mac.

La herramienta es de fácil uso, captura varios niveles de vínculos dentro del sitio, tiene un sello de tiempo que estampa fecha y hora de las páginas Web capturadas, y Adobe asegura compatibilidad con versiones anteriores. Sin embargo, no hay metadatos capturados; reproduce un documento plano en PDF, lo que significa que no es posible remover una porción de una página para imprimir, por ejemplo sólo una fotografía, por lo que se tiene que imprimir toda la página; el sitio Web completo es capturado cada vez, y la herramienta convierte el sitio Web a PDF simple en lugar de PDF/A.

Adobe liberó la herramienta de captura de sitios Web en 2008 y es una herramienta extremadamente simple para implementar y usar. Adobe tiene una buena reputación y una historia de soporte para el cliente, y trata de asegurar que cada producto nuevo sea compatible retrospectivamente con varias versiones previas.

Una herramienta más para archivar sitios Web (construidos dinámicamente con una base de datos subyacente para hospedar la información que es generada por el usuario sobre la marcha), es la de archivar la base de datos. La técnica es apenas incipiente pero vale la pena describirla con cierto detalle ya que es una herramienta que puede ser utilizada para mitigar problemas asociados con el archivado de sitios Web dinámicos mediante el uso de métodos de sitio Web estáticos.

Brown describe el proceso de archivar sitios Web por medio de bases de datos en tres etapas: Primero el repositorio define el modelo estándar de datos y formato para las bases

³⁷ Véase el sitio Web de *HTTrack* para mayor información en: <http://www.httrack.com/>

de datos a archivar; luego la base de datos fuente es convertida al formato estándar; y finalmente, se proporciona una interfaz de acceso estándar para las bases de datos archivadas.³⁸

Los Archivos Federales de Suiza han desarrollado un formato basado en XML que permite la preservación en el largo plazo de contenidos de bases de datos relacionales de manera independiente de los programas o aplicaciones. El formato tiene una larga historia de desarrollo desde principio de los noventas. En mayo de 2008 fue aceptado como el formato del proyecto europeo PLANETS para archivar bases de datos relacionales. El formato se conoce como SIARD por sus siglas en inglés - “*Software Independent Archiving of Relational Databases*”. Preserva datos de contenido y metadatos así como las relaciones en un formato que cumple con normas ISO. Un artículo informativo publicado en octubre de 2008 por la organización Digital Preservation Europe, “*Database Preservation: The International Challenge and the Swiss Solution*” describe el proceso de SIARD.³⁹

Un archivo SIARD es un contenedor ZIP estructurado no comprimido (ZIP-64 estándar), que permite prácticamente cualquier tamaño de archivo. Contiene dos carpetas: “encabezado” y “contenido”. La carpeta *encabezado* almacena el contexto de la base de datos y metadatos. Un sólo archivo: *metadata.xml* asegura que podamos entender los antecedentes técnicos así como los de contexto de la base de datos. En términos técnicos, SIARD registra el nivel más alto (la base de datos) el identificador, la versión de formato, el código de resumen de mensaje de la terminal PC que archiva (verifica los datos primarios de integridad), etc. Sobre el nivel de esquema, SIARD almacena listas de tablas, vistas y rutinas. A nivel de tabla, SIARD registra las limitantes y los “disparadores” de eventos. Más a fondo, en el nivel de columna, SIARD también especifica el tipo de lenguaje SQL en uso, los nombres de objetos de gran volumen, y lo más importante de todo: claves externas y claves candidatas con datos inter-referenciales; es decir, las relaciones. Al mismo tiempo SIARD contextualiza los datos. En el nivel de base de datos nos permite registrar o aumentar (con la suite SIARD) información sobre la procedencia del archivo, descripción, usuario, etc. En niveles más bajos nos permite mantener detalles de las tablas y nombres de las columnas y el contenido. Toda esta información descriptiva hace que la base de datos sea comprensible para los usuarios futuros, tanto en términos contextuales como técnicos.

³⁸ Brown, “*Archiving Websites*”, p. 59.

³⁹ Según el artículo informativo publicado en octubre 2008 por Digital Preservation Europe, “*Database Preservation: The International Challenge and the Swiss Solution*” (http://www.digitalpreservationeurope.eu/publications/briefs/database_preservation.pdf), “*El uso de estándares ISO extensamente aceptados asegura en gran medida que los datos almacenados puedan ser accedidos en el futuro. Basados en esta presunción los documentos de archivo SIARD, tanto datos como metadatos se recaban automáticamente en formatos de la norma ISO: SQL 1999 UNICODE, y el más importante de éstos: XML1.0. Para asegurar la normalización SIARD convierte todas las bases de datos propietarias en el equivalente al conjunto de caracteres UNICODE. Además SIARD no archiva sinónimos ya que no son parte del SQL:1999 estandarizado. Apegarse a los estándares es una regla de hierro.*”

La segunda carpeta,- la de contenido-, almacena los datos primarios. Los datos son archivados de acuerdo a la estructura de la base de datos. Para cada esquema SIARD genera automáticamente una carpeta (cuadro 1, cuadro 2, etc.) que contiene la serie de tablas correspondientes como subcarpetas (tabla 1, tabla 2, etc.). Los datos mismos son almacenados en archivos XML (por ej. Tabla1.xml). Esta definición de esquema refleja el esquema de metadatos de la tabla SQL, y especifica que la tabla es almacenada como una cadena de líneas que abarcan una secuencia de entradas de columna con tipos XML diferentes. Los grandes conjuntos u objetos binarios o de caracteres (BLOB's y CLOB's que contienen todos los tipos de información) también son archivados. Ellos son almacenados en carpetas generadas automáticamente (por ej., lob1, lob2, etc.) ya sea en archivos TXT o BIN (record1.text, o record1.bin, etc.).⁴⁰

SIARD es también un formato abierto, lo cual significaría que la organización que colecciona podría de hecho archivar la base de datos sin costos adicionales posibles por obtener una licencia al propietario del sistema administrador de contenidos requerido si el método de transferencia directa de captura es empleado.

De la descripción anterior de SIARD, está claro que la institución que colecciona necesitará contar de inicio con una persona con preparación tecnológica para implementar exitosamente la “suite” o ensamble de programas SIARD.

En estos momentos es incierto si el ensamble SIARD está disponible actualmente para uso público. En la presentación de Jean-Marc Comment, representante de los Archivos Federales de Suiza, durante el 16º Congreso Internacional de Archivos en Julio de 2008⁴¹ se mencionó que las herramientas de SIARD estarían disponibles en un futuro próximo por parte de los Archivos Federales de Suiza. Hasta el 12 de Octubre 2009 en el Sitio Web⁴² no aparecía nada respecto de las herramientas SIARD. Debido a la duda sobre la disponibilidad, la Suite SIARD no ha sido incluida como una opción de preservación en este reporte. Sin embargo, es una opción interesante que puede ser seguida una vez que su disponibilidad sea cierta.

4 Plan de Acción general para la Preservación de un Sitio Web

Aunque no existe una solución genérica para los planes de preservación de sitios Web, existen ciertos elementos que son universales para todos los planes y/o programas. Cuando la estrategia más apropiada ha sido identificada, debe ser seleccionado un equipo que incluya gestores de documentos o archivistas, administradores de sitios Web, administradores de comunicación y personal de tecnologías de la información. El equipo puede desarrollar un plan de acción general que incluya políticas y procedimientos adecuados a sus necesidades.

⁴⁰ Para una discusión más completa del formato SIARD, véase “*SIARD Format Description*”, disponible para descarga en: <http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=de>

⁴¹ Para ver toda la presentación, véase: http://www.planets-project.eu/docs/presentations/ICA2008_Comment_SIARD.pdf

⁴² Sitio Web de los Archivos Federales de Suiza: <http://www.bar.admin.ch/index.html?lang=en>

A continuación se presenta un plan de acción general para la preservación de un sitio Web que puede ser adaptado a las diferentes necesidades de una institución:

- 1) Identificar los requisitos de gestión archivística para la actividad basada en Web.
- 2) Determinar si el sistema existente satisface los requisitos señalados o si es necesario diseñar e implementar un nuevo sistema o mejorar el sistema actual.
- 3) Elevar el perfil y la conciencia general dentro de la organización de las responsabilidades de gestión archivística de todo el personal.
- 4) Llevar a cabo una evaluación de riesgos para determinar el nivel de riesgo aceptable planteado.
- 5) Desarrollar una política global para la preservación de sitio Web (o una política de preservación de documentos de archivo que incluya la preservación del sitio Web).
 - a) Desarrollar una Política de Recopilación (incluye una política de selección)⁴³
 - i) Desarrollar un política de selección
 - ii) Definición del contexto
 - iii) Métodos de selección
 - iv) Criterios de selección
 - (1) Valoración de contenido
 - (2) Extensión⁴⁴
 - v) Lista de recolección
 - (1) Selección de los recursos del sitio Web a recuperar
 - vi) Definiciones de límites
 - (1) Determinar URL o nombre de dominio⁴⁵
 - (2) Parámetros⁴⁶
 - vii) Definir el método de recolección
 - viii) Determinar tiempos y frecuencia de recolección, -incluyendo la metodología de evaluación de riesgos⁴⁷

⁴³ Tanto la política como las listas de recolección deberán ser revisadas periódicamente para aumentar o quitar recursos.

⁴⁴ Es necesario establecer criterios para determinar la extensión de los recursos seleccionados; -esto es, si se coleccionarán o no vínculos externos.

⁴⁵ Si un solo recurso de Web, tal como una página o documento va a ser coleccionado de forma aislada, entonces la lista de recolección simplemente necesita especificar el URL de ese recurso. Si todo el sitio Web ha sido seleccionado, usualmente será definido por el nombre de dominio.

⁴⁶ Los parámetros definen el número de niveles de la estructura de directorio que será coleccionada y si los vínculos externos serán o no seguidos y si así fuera a qué profundidad.

⁴⁷ La Universidad de Cornell ha desarrollado una metodología para evaluar y mitigar riesgos para recursos de Web vivos: El Proyecto VRC, -Virtual Remote Control- de Cornell, está disponible en: Resumen: La metodología VRC para el análisis de riesgos sigue un proceso de seis pasos que inicia con la identificación y la evaluación de sitios Web, facilita la evaluación del nivel de riesgos del sitio y la estrategia de construcción, e inicia una respuesta subsecuente. El catálogo VRC busca automatizar este proceso lo más posible, pero permite el control humano. La estabilidad del sitio Web es medida a varios niveles de riesgo que pueden ser deducidos mediante la supervisión conforme avanza el tiempo respecto de su

- (1) Influencia del ciclo de vida del recurso de Web
- (2) Tasa o rango de cambio de contenido
- (3) Actualidad y significatividad
- ix) Definir almacenamiento para los activos digitales⁴⁸
- 6) Implementar la política
- 7) Documentar procedimientos y procesos para asegurar que las estrategias se llevan a cabo.
- 8) Iniciar el programa de preservación del sitio Web.
- 9) Realizar verificaciones en los datos capturados y almacenados.⁴⁹
- 10) Revisar la política y los objetivos de valoración de forma frecuente.

5 Estudio de Caso: Desarrollo de un plan de preservación de un sitio Web para una asociación académica en una institución académica.

Esta sección analiza un estudio de caso que involucró el desarrollo de un plan de preservación de un sitio Web hipotético para una asociación de estudiantes dentro de una institución académica grande. Este estudio ofrece un ejemplo sobre cómo identificar requisitos y desarrollar un plan de preservación de un sitio Web organizacional. La meta del estudio fue desarrollar estrategias para ejercer un mayor control de las modificaciones del sitio Web de la asociación y para la preservación en el largo plazo de varias iteraciones al paso del tiempo.

5.1 Antecedentes sobre la Organización.

La Asociación de Estudiantes de la Universidad Acme (ASUA)⁵⁰ está ubicada en el campus de la Universidad Acme y es una asociación de estudiantes de la Universidad. La asociación consiste en 30,000 miembros compuesta por estudiantes del campus principal y de campi satélites de la Universidad. La ASUA opera como una asociación

implementación (por ej., arreglo u orden de HTML) y su estructura de hipervínculo, así como desde los metadatos del servidor de Web (p. ej., programas del servidor, tiempo de respuesta). Si el sitio Web se encuentra en un riesgo alto puede ser necesario establecer contacto con el dueño del sitio. VRC planea entonces establecer recomendaciones en los *'lineamientos para la preservabilidad de contenidos de Web'*. Como último recurso para recursos de Web en riesgo éstos pueden ser también cosechados y preservados para evitar su pérdida. Resumen de ERPANET,

⁴⁸ Para asegurar la accesibilidad en el largo plazo, es esencial que los medios de almacenamiento sean refrescados periódicamente; la acción de refrescar el medio de almacenamiento deberá ser construida dentro de la política general de documentos de archivo electrónicos como se vio en los pasos de los planes de acción.

⁴⁹ Una vez que el sitio Web ha sido capturado y transferido al ambiente del archivo histórico, se deben llevar a cabo verificaciones para asegurar que todas las partes del sitio Web capturado está trabajando como deberían. Las verificaciones incluyen, entre otras: teclear manualmente a través de todos los hipervínculos; teclear aleatoriamente algunos hipervínculos, o emplear el uso de una aplicación de prueba de existencia de vínculos como ayuda a automatizar el proceso de verificación. Ejemplos de aplicaciones de prueba de vínculos son: Link Checker Pro: <http://www.link-checker-pro.com/> Site Audit:

http://www.blossom.com/site_audit.html Cyber Spyder Link Test: <http://www.cyberspyder.com/cslnkts1.html> y Link Sleuth: <http://home.snafu.de/tilman/xenulink.html>.

⁵⁰ Una universidad y una sociedad ficticias.

independiente sin fines de lucro. El propósito de la ASUA es supervisar los servicios a estudiantes (p. ej., tutorías, búsqueda de empleos, etc.), negocios y clubs. La ASUA es el centro de gestión archivística y archivo histórico para la Sociedad “Alma Mater”.

El archivista de ASUA buscó estrategias para la preservación en el largo plazo de un sitio Web que cambia frecuentemente. El archivista estuvo interesado en desarrollar estrategias para ejercer un mayor control sobre las modificaciones al sitio Web y para la preservación en el largo plazo de iteraciones al paso del tiempo. El resultado del estudio es un plan de acción que concibe estrategias para el control sobre el sitio Web y su preservación en el largo plazo.

5.2 Los Retos

La ASUA tiene recursos limitados (experiencia tecnológica, personal, tiempo, recursos financieros, etc.) para desarrollar y sustentar una estrategia para la preservación en el largo plazo de su sitio Web.

5.3 El Proceso de Desarrollo

El sitio Web de ASUA fue identificado como el cuerpo de material digital para el cual un plan de preservación sería desarrollado. Los datos coleccionados fueron acerca del contexto de la institución y sus limitantes, el cuerpo específico de material, sus formas documentales, las restricciones tecnológicas y el sentido funcional y cultural de los materiales.

Identificar el Contexto

Se compiló información relacionada con la institución, sus documentos de archivo y sus operaciones a través de una aproximación etnográfica. Se llevaron a cabo varias entrevistas y observaciones con el archivista de la Asociación, el Gerente de Comunicaciones, el Editor del Sitio Web y el Administrador de Tecnologías de la Información, para obtener como resultado el análisis contextual y diplomático así como información de los documentos de archivo producidos por la ASUA y una perspectiva cultural de los responsables del sitio Web.

Procedimiento para Desarrollar un Procedimiento de Mantenimiento

Se emprendió el desarrollo de un documento de procedimientos que delinea cómo el sitio Web de AUSA va a mantenerse. Este documento establece los procedimientos para el mantenimiento del sitio Web y los procedimientos que contienen los criterios que se seguirán para lo que puede residir en el sitio Web. Tales procedimientos gobernarán el contenido del sitio Web tomando en consideración las restricciones que pueden ser establecidas en el contenido, en parte debido a la necesidad de adherirse a requisitos y normas administrativas, jurídicas y legales. Contar con tales procedimientos también asegura una valoración precisa y exhaustiva en el futuro.

Valorar el sitio Web

Se llevó a cabo la valoración del sitio Web de ASUA para establecer lo que sería preservado. Se hicieron cuatro preguntas clave: 1) ¿Qué capturar? 2) ¿Con qué frecuencia capturar? 3) ¿Qué tanto capturar? y 4) ¿Por cuánto tiempo preservar lo que se captura?

Investigación sobre la Mejor Estrategia para preservar el Contenido del Sitio Web

Investigar e identificar las opciones tecnológicas que cumplan los objetivos de valoración de ASUA y sus restricciones tecnológicas, financieras y humanas; identificar los costos de los recursos para implementar las opciones tecnológicas identificadas. La investigación buscó identificar métodos para la preservación del sitio Web que han sido instrumentados exitosamente en otras organizaciones parecidas, así como averiguar sobre el conocimiento construido por grandes organizaciones. Además se investigó sobre los métodos que no han sido instrumentados actualmente.

Muchas grandes organizaciones han contribuido en el desarrollo de métodos para la captura y preservación del sitio Web, y las mismas fueron investigadas en cuanto a metodologías probadas. Dentro de las grandes organizaciones que actualmente preservan sitios Web útiles para este proyecto se tienen a la Biblioteca del Congreso, el Archivo Internet (Internet Archive), los Archivos Nacionales del Reino Unido y los Archivos Nacionales de Australia. Cada una de estas instituciones fue de gran ayuda para desarrollar un entendimiento sobre los componentes que se requieren integrar en una estrategia de preservación. La mayoría de la información es fácilmente adaptable a las necesidades de organizaciones de tamaño mediano y pequeño y sin esta información muchas instituciones pequeñas no podrían emprender tales programas de preservación.

El *Archivo Internet* ha venido desarrollando soluciones de código abierto para operaciones de cosecha remota que no requieren de recurso monetario, pero sí requieren de un conocimiento tecnológico aceptable. Los Archivos Nacionales del Reino Unido han llevado a cabo investigación para el mejor medio de almacenamiento, elaboraron una guía simple para archivar sitios Web, y también han investigado sobre los formatos de archivo óptimos para la creación de datos. Los Archivos Nacionales de Australia han desarrollado investigaciones sobre requisitos de metadatos que son clave para manejar de forma efectiva todos los documentos de archivo, incluyendo aquellos basados en la actividad soportada en Web. También han investigado sobre soluciones para registrar evidencia de documentos de archivo basados en Web en sitios que cambian frecuentemente cuando existen rastreadores poco frecuentes. La Biblioteca del Congreso también ha llevado a cabo investigación sobre metadatos específicamente para preservación (PREMIS es un diccionario de datos y de soporte de esquemas XML para los metadatos medulares que sustenten la preservación de materiales digitales en el largo plazo), y también realizaron investigación para desarrollar otros esquemas de metadatos (METS -por sus siglas en inglés-, Estándar para Codificación y Trasmisión de Metadatos es una estructura de metadatos para metadatos de codificación, descriptivos, administrativos y estructurales que produce los Auxiliares de Búsqueda Descriptivos de Codificación Archivística).

Desarrollar un Plan de Acción para la Preservación del Sitio Web⁵¹

Se desarrolló un plan que incluye estrategias, protocolos, requisitos, procedimientos y resultados esperados.

6 Preguntas de Revisión

- 1) Nombre entre tres a seis criterios que deberían ser considerados cuando se seleccionan formatos de archivo para la producción de datos.
- 2) ¿Cuáles son los dos tipos de sitios Web y de contenidos de sitios Web y cómo difieren?
- 3) Nombre y describa cada uno de los tres métodos de captura de sitios Web y analice cómo difieren.
- 4) Nombre tres grandes organizaciones que tienen buenos recursos para archivar sitios Web e identifique su experiencia.
- 5) ¿Cuándo se deben llevar a cabo verificaciones de validación?
- 6) Identifique tres factores que afectan la longevidad de los medios electrónicos.
- 7) ¿Cuáles son los diez principales pasos en el *Plan General de Acción para la Preservación de Sitios Web*?

⁵¹ Véase la sección 3.0 para el *Plan General de Acción para la Preservación de Sitios Web* que puede ser adaptado para uso organizativo.

7 Recursos Adicionales

Autor: Brown, Adrian

Título: “*Archiving Websites*”

Fecha de Publicación: 2006

Fuente/Editor: London: Facet Publishing

Este libro es un texto amplio que ofrece guía práctica sobre programas para archivar sitios Web. Se trata de una combinación de mejores prácticas, consejos prácticos y guía para establecer un programa para archivar sitios Web; desde los aspectos legales y los métodos de recolección al programa de gestión y una mirada a las tendencias futuras. Este libro es un recurso valioso para los profesionales de la información y gestión archivística que buscan desarrollar un plan o programa para archivar sitios Web.

Autor: Kenney, Anne R., Nancy Y. McGovern, Peter Botticelli, Richard Entlich, Carl Lagoze and Sandra Payette

Título: “*Preservation Risk Management for Web Resources. Virtual Remote Control in Cornell’s Project Prism*” (Administración de Riesgos de la Preservación de Recursos de Web. Control Virtual Remoto en el Proyecto Prism de Cornell)

Fecha de Publicación: 2002

Fuente: D-Lib Magazine 8(1)

URL: <http://www.dlib.org/dlib/january02/kenney/01kenney.html>

Autor: Library of Congress (Biblioteca del Congreso de los E.U.A.)

URL: <http://www.digitalpreservation.gov/>

La Biblioteca del Congreso de los E.U.A. tiene una variedad de excelentes programas y recursos que ofrecen información y ejemplos sobre requisitos para la preservación digital, investigación y mejores prácticas que incluyen la “National Digital Stewardship Alliance” o Alianza Nacional de Administración Digital, El “*Digital Preservation Outreach*” o Participación y Educación de la Preservación Digital y el “*Education and the National Digital Information Infrastructure and Preservation Program*” o Programa de Preservación e Infraestructura de la Información Digital Nacional.

Autor: National Archives of Australia

Título: “*Archiving Web Resources: Guidelines for Keeping Records of Web-based Activity in the Commonwealth Government*” (Archivando Recursos de Web:

Lineamientos para Gestionar Documentos de Archivo de la Actividad basada en Web en el Gobierno de la *Commonwealth*)

Fecha de Publicación: Marzo, 2001

URL http://www.naa.gov.au/Images/archweb_guide_tcm16-47165.pdf

Autor: Shepherd, Elizabeth y Geoffrey Yeo

Título: “*Managing Records: A Handbook of Principles and Practice*” (Gestión de Documentos de Archivo: Un Manual de Principios y Prácticas)

Fecha de Publicación: 2003

Editor: London: Facet Publishing

Este libro es un texto amplio que delinea los principios de la gestión de documentos de archivo y su implementación práctica en organizaciones. Es exhaustivo en su cobertura respecto de los conceptos de gestión archivística. Los temas incluyen: contexto organizacional; clasificación, producción y captura de documentos de archivo, valoración, retención y disposición, acceso e implementación. El libro incluye una bibliografía amplia de recursos para la gestión archivística así como listas de estándares nacionales e internacionales sobre gestión archivística y organizaciones profesionales para gestores de documentos y archivistas.

Autor: The Internet Memory Foundation

URL: <http://internetmemory.org/en/>

La Fundación Memoria de Internet es una institución sin fines de lucro que apoya activamente la preservación de Internet como medio para propósitos patrimoniales y culturales.

Autor: The Internet Archive

URL: <http://internetmemory.org/en/>

El Archivo Internet desarrolló “Archive-It” (analizado en este módulo) el cual tiene archivadas cerca de 150 millones de páginas Web desde 1996. También hospeda un blog sobre el desarrollo del trabajo en curso de la *Máquina para Remontar el Tiempo*. (Wayback Machine): <http://iaWebarchiving.wordpress.com/>

Autor: The National Archives UK

Título: “*Preservation Risk Management for Web Resources. Virtual Remote Control in Cornell’s Project Prism*” (Archivado de Recursos Web: Lineamientos para Gestionar Documentos de Archivo de la Actividad basada en Web en el Gobierno de la *Commonwealth*)

URL: <http://www.nationalarchives.gov.uk/news/734.htm>

Los Archivos Nacionales del Reino Unido son una fuente valiosa de información sobre la captura de sitios Web, particularmente para capturar sitios Web del gobierno del Reino Unido.

7.0 Referencias Bibliográficas

- Herramienta de Adobe para captura de Web. Información de producto disponible en:
<http://ww.adobe.com/products/acrobat/>
- Blue Squirrel, “*Grab-a-Site Product Page*”. Disponible en:
<http://www.bluesquirrel.com/products/grabasite>
- Brown, Adrian. *Archiving Websites* (London: Facet Publishing, 2006).
- Brown, Adrian (Agosto, 2008), “*Digital Preservation Guidance Note 2: Selecting Storage Media for Digital Preservation.*” Disponible en:
<http://www.nationalarchives.gov.uk/documents/selecting-storage-media.pdf>
- Brown, Adrian, “*Selecting File Formats.*” Disponible en:
<http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>
- Canadian Conservation Institute, “*Electronic Media Collections Care for Small Museums and Archives*”. Disponible en: <http://www.cci-icc.gc.ca/caringfor-prendresoindes/articles/elecmediacare/index-eng.aspx>
- Digital Preservation Europe (Octubre 2008), “*Database Preservation: The International Challenge and the Swiss Solution*”. Disponible en:
http://www.digitalpreservationeurope.eu/publications/briefs/database_preservation.pdf
- European Electronic Resource Preservation and Access Network (ERPANET), “*Digital Preservation Policy Tool*”, 2003. Disponible en:
<http://www.erpanet.org/guidance/docs/ERPANETPolicyTool.pdf>
- Greenwood, David J. and Morten Levin, “*Reconstructing the Relationships between Universities and Society through Action Research*”. En: Norman K. Denzin and Yvonna S. Lincoln, eds., “*The Landscape of Qualitative Research: Theories and Issues*” 2nd ed. (Thousand Oaks: SAGE Publications, 2003), 131-166.
- Harvey, Ross, “*Preserving Digital Materials*”. (Munich, Germany: K. G. Saur, 2005).
- HTTrack Website. Disponible en: <http://www.httrack.com/>
- Internet Archive, Heritrix Website. Disponible en: <http://crawler.archive.org>
- InterPARES 3 Project, “*Case Study 09 Alma Mater Society of the University of British Columbia*”. Final Report.
- ISO 19005-1:2005 “*Document Management - Electronic document file format for long term preservation - Part 1: Use of PDF 1.4 (PDF/A-1).*”
- Jones, Maggie and Neil Beagrie, “*Preservation Management of Digital Materials : A Handbook*” (London, UK: The British Library, 2001).
- Kenney, Anne R., Nancy Y. McGovern, Peter Botticelli, Richard Entlich, Carl Lagoze y Sandra Payette (2002) “*Preservation Risk Management for Web Resources. Virtual Remote Control in Cornell’s Project Prism,*” En: D-Lib Magazine 8(1). Disponible en:
<http://www.dlib.org/dlib/january02/kenney/01kenney.html>

Lazinger, Susan S. *“Digital Preservation and Metadata. History, Theory, Practice”* (Englewood, CO: Libraries Unlimited, 2001).

Library of Congress (United States). Disponible en: <http://www.digitalpreservation.gov/>

Linden, Jim, Sean Martin, Richard Masters y Roderic Parker, *“The Large-Scale Archival Storage of Digital Objects,”* En: *DPC Technology Watch Series Report 04-04*, Digital Preservation Coalition (February 2005). Disponible en: <http://www.dpconline.org/docs/dpctw04-03.pdf>

McGovern, Nancy, Anne R. Kenney, Richard Entlich, William R. Kehoe y Ellie Buckley (2004), *“Virtual Remote Control. Building a Preservation Risk Management Toolbox for Web Resources,”* En: *D-Lib Magazine* 10(4). Disponible en: <http://www.dlib.org/dlib/april04/mcgovern/04mcgovern.html>

National Archives of Australia, *“Archiving Web Resources: Guidelines for Keeping Records of Web-based Activity in the Commonwealth Government,”* (March 2001). Disponible en: http://www.naa.gov.au/Images/archweb_guide_tcm16-47165.pdf

National Archives of Australia. (2004). *“Digital recordkeeping self-assessment checklist”*.

Pantalony, Rina Elster. *“Protecting your Interests: A legal guide to negotiating Website development and virtual Exhibition Agreements”*. (Ottawa, Canada: Minister of Public Works and Governments Services Canada, 1999).

Peters McLellan, Evelyn, *“General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation,”* InterPARES 2 Project (Marzo 2007). Disponible en: http://www.interpares.org/display_file.cfm?doc=ip2_gs11_final_report_english.pdf

Prom, Christopher J. y Ellen D. Swain (2007). *“From the College Democrats to the Falling Illini: Identifying, Appraising, and Capturing Student Organization Websites,”* En: *American Archivist* 70: 344-363.

Smiraglia, Richard P. *“Metadata : A Cataloger’s Primer”*. (New York, NY: The Hawthorn Press, 2005).

Society of American Archivists. Conference / Workshop Calendar 2009.

Swiss Federal Archives, *“SIARD Format Description.”* Disponible en: <http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=de>