



World Library and Information Congress: 70th IFLA General Conference and Council

22-27 August 2004
Buenos Aires, Argentina

Programme: <http://www.ifla.org/IV/ifla70/prog04.htm>

Code Number: 026-S
Meeting: 148. IFLA-CDNL Alliance for Bibliographic Standards (ICABS)
Simultaneous Interpretation: -

**Coleccionar y gestionar recursos web para su acceso a largo plazo:
recolección web y directrices para apoyar la preservación (Acciones 3.3 y 3.4
de ICABS)**

Pam Gatenby

Asistente al Director General
Gestión de la Colección
Biblioteca Nacional de Australia
pgatenby@nla.gov.au

Resumen:

La Biblioteca Nacional de Australia contribuye al plan de trabajo de ICABS a través de la Línea de Acción 3 – avanzar en el conocimiento de las cuestiones relacionadas con el archivo a largo plazo de los recursos digitales. La Biblioteca Nacional se comprometió al desarrollo de acciones para promover la colaboración en las áreas del archivo web (Acción 3.3.) y la preservación de recursos digitales (Acción 3.4) sirviéndose de sus propias experiencias en este campo. Las acciones específicas incluyen dirigir el Grupo de Trabajo de Internet Invisible (Deep Web Working Group) del International Internet Preservation Consortium (IIPC), acoger la celebración de una conferencia internacional sobre el archivo web, trabajar en la automatización del depósito y del archivo de las publicaciones gubernamentales en línea, promover la información sobre los métodos para archivar la web, y examinar la información disponible para guiar la toma de decisiones sobre preservación digital. El objetivo es facilitar la disponibilidad de información referida a estas áreas de interés, y hacer la información fácilmente accesible a través del portal PADI (<http://www.nla.gov.au/padi/>).

1. Introducción

La Biblioteca Nacional de Australia se complace en ser uno de los participantes en la alianza ICABS comprometiéndose a colaborar en el desarrollo de estándares y

estrategias dirigidas a fomentar el coste-efectividad del acceso a las colecciones patrimoniales nacionales.

Nuestra mayor contribución al programa ICABS se centra en la Línea de Acción 3 – Avanzar en el conocimiento de las cuestiones relacionadas con el archivo a largo plazo de los recursos digitales.

Hemos elegido contribuir en esta Línea de Acción en concreto porque tenemos experiencia en el archivo digital y porque somos conscientes de que muchas Bibliotecas Nacionales solicitan asesoramiento en esta área porque se encuentran inseguros sobre la manera de proceder. Además, reconocemos que, para continuar con nuestro propio desarrollo, es necesario trabajar en colaboración práctica con otras instituciones dado que existen demasiados temas como para poder resolverlos de manera individual.

ICABS nos ha animado a pensar cuidadosamente como podemos compartir nuestro conocimiento y experiencias de modo mucho más efectivo y como podemos contribuir más convenientemente al objetivo del acceso a largo plazo a los recursos digitales de importancia nacional.

2. Antecedentes del archivo digital y la actividad de preservación en la Biblioteca Nacional Australiana

Archivo digital

La Biblioteca Nacional de Australia ha estado recogiendo importantes sitios web australianos y gestionándolos en PANDORA: el archivo web de Australia desde 1996 (<http://www.nla.gov.au/pandora/>). Hoy en día, la actividad se centra en el desarrollo rutinario de la colección llevado a cabo por un equipo de 5 personas. Los títulos se recogen con el permiso de sus editores ya que la legislación Commonwealth de Australia sobre depósito legal aún no se aplica a las publicaciones electrónicas.

Actualmente, el archivo incluye alrededor de 6000 títulos y más de 10000 referencias (o instancias) de sitios web y su tamaño supera la mitad de un terabyte. El desarrollo de PANDORA es una tarea realizada en colaboración por las bibliotecas estatales australianas y otras tres instituciones nacionales que se dedican a reunir documentación. La selección de sitios se basa en criterios detallados que ponen el énfasis en la investigación y el valor de la información al igual que su importancia cultural, e incluso ofrecen el ámbito de actuación para la selección temática. Cada sitio web archivado es evaluado en función de su contenido y funcionalidad.

Para gestionar y apoyar sus responsabilidades en el archivo digital, la Biblioteca desarrolló su propio Sistema de Archivo Digital llamado PANDAS (Pandora Digital Archiving System). El sistema permite las siguientes funciones:

- gestionar metadatos sobre títulos seleccionados y rechazados
- iniciar la recogida de los títulos seleccionados
- gestionar la calidad en el proceso de revisión
- preparar ejemplares para su visualización pública en el Archivo
- gestionar la restricción de accesos
- gestionar la identificación permanente

- proporcionar informes de gestión

Varias instituciones que recogen información a lo largo del mundo han preguntado sobre el uso de PANDAS ya que no existen aún sistemas comerciales similares que permitan coleccionar y gestionar recursos web. En respuesta a sus preguntas, la Biblioteca ha hecho a PANDAS accesible a consultas con propósitos de evaluación. También estamos dispuestos a poner el software a disposición de otros para su implantación, entendiendo que nosotros no podremos ofrecer servicios de soporte.

Preservación Digital

Las consideraciones y estrategias sobre preservación han sido integrales desde el principio del desarrollo de las colecciones digitales de la Biblioteca. El objetivo de la recogida de recursos web que son archivados en PANDORA es gestionarlos a lo largo del tiempo con la intención de proporcionar acceso público continuo a ellos. Esto requiere no solo una planificación a corto y largo plazo sino también una gestión diaria de los recursos archivados. Una Política de Preservación Digital (<http://www.nla.gov.au/policy/digpres.html>) y un plan de acción describen las directrices y actividades clave que la Biblioteca persigue. El énfasis se da a:

- gestión de datos (actualizaciones, copias de seguridad, etc)
- identificar y documentar las características de los recursos archivados
- identificar y gestionar los riesgos y amenazas asociados con los diferentes formatos de archivos
- desarrollar estrategias de preservación para cada formato de archivo
- usar metadatos de preservación digital
- planes de contingencia

La colaboración es también básica para la metodología de la Biblioteca en cuanto a preservación digital. Para facilitar la colaboración, la Biblioteca mantiene el portal PADI (Preserving Access to Digital Information) (<http://www.nla.gov.au/padi>). PADI ofrece enlaces a un amplio espectro de información relacionada con la preservación digital ofreciendo así mismo información contextual sobre los temas clave. La posibilidad de maximizar la información compartida es vital para nuestra aportación a ICABS.

3. La participación de la Biblioteca Nacional en ICABS

La Biblioteca Nacional de Australia colabora en el Programa de Acción ICABS en dos áreas de actuación: Recogida de webs (acción 3.3.) y Preservación de materiales digitales (acción 3.4)

Recolección web (Acción 3.3 de ICABS)

El objetivo de nuestras actividades en esta área es explorar y promover metodologías para recoger y archivar publicaciones web e identificar los temas asociados. El propósito es hacer esta información ampliamente disponible y fácilmente accesible a través de PADI con la intención concreta de asistir a aquellas instituciones que están planeando o están revisando sus programas de archivo y preservación digital.

La actividades clave se dirigen a:

- Dirigir el trabajo del Grupo de Trabajo de Internet Invisible (Deep Web) del International Internet Preservation Consortium (IIPC)
- Acoger una conferencia internacional sobre el archivo de recursos web
- Automatizar el depósito y archivo de la recursos gubernamentales en línea
- Explorar y promover métodos para archivar las publicaciones basadas en web obtenidas por medio de la recolección web

(i) Dirigir el trabajo del Grupo de Trabajo de Internet Invisible

El International Internet Preservation Consortium (IIPC) se creó en 2003 con la intención de promover la colaboración entre Bibliotecas Nacionales para el desarrollo de estándares y herramientas que apoyen el archivo y preservación de los recursos web. Once Bibliotecas Nacionales junto con el Internet Archive son miembros del consorcio bajo la dirección de la Biblioteca Nacional de Francia. El enfoque del consorcio es práctico. Las Bibliotecas participantes abonan una cuota de suscripción y están obligadas a contribuir activamente en los programas de trabajo del grupo.

La Biblioteca Nacional de Australia lidera el Grupo de Trabajo del Archivo de la Internet Invisible (Deep Web Archiving Working Group) el cual investiga la identificación, adquisición, almacenamiento y visualización de las publicaciones y sitios web que están gestionados por una base de datos. Esta categoría de recursos web es un tema de interés para nosotros ya que un número cada día mayor de recursos importantes son suministrados de esta manera. Por razones técnicas, actualmente no tenemos posibilidad de recoger este tipo de recursos.

Los objetivos del Grupo de Internet Invisible son:

- identificar la información requerida de los editores para las bases de datos depositadas.
- desarrollar un modelo de datos y esquema para archivar los contenidos de las bases de datos en formato XML; y
- desarrollar una herramienta de acceso en línea para consultar y navegar por bases de datos estructuradas archivadas.

El software desarrollado estará alojado en un dominio público y estará disponible para ser usado por cualquiera que esté interesado en él.

Otros Grupos de Trabajo de IIPC trabajan en temas como la definición de los requerimientos de las arquitecturas para archivos; la identificación de las necesidades de los investigadores en el uso de archivos web y el desarrollo de herramientas de acceso y un software de rastreo web (*web crawling*).

(ii) Conferencia Internacional sobre el archivo de recursos web
(<http://www.nla.gov.au/webarchiving/>)

La Biblioteca Nacional en Camberra acogerá los próximos días 9 al 11 de Noviembre la celebración de una conferencia internacional sobre el archivo de recursos web y los

problemas de las instituciones que se dedican a recogerlos. Los siguientes actos tendrán lugar junto con la conferencia:

- una reunión del Comité Directivo del IIPC
- el 8 de Noviembre se celebra el Día de Australia donde se presentarán diversos proyectos australianos relacionados con los recursos digitales en general
- el 12 de Noviembre se celebra el Día de la Información donde se presentarán herramientas y metodologías de archivo web procedentes de todo el mundo.

El objetivo principal de la conferencia es identificar los temas más importantes que afectan a las instituciones relacionadas con el patrimonio cultural en la recogida de recursos web, y explorar cómo estos problemas están siendo tratados internacionalmente. Se dará un énfasis especial a revisar las metodologías que están aplicando las instituciones que se dedican a su recolección y a la identificación de proyectos de investigación en curso para apoyar los diferentes enfoques. La conferencia no pretende ser técnica, sino concentrarse en las necesidades del trabajo y estrategias de las instituciones que se dedican a la recolecta.

La información reunida durante la conferencia, que sea relevante para el ICABS en relación con sus actividades sobre el archivo web, se empleará para incrementar la información que actualmente está disponible a través de PADI.

(iii) Depósito y archivo automatizado de los recursos gubernamentales en línea

A principios de 2003 la Biblioteca puso en marcha el Commonwealth Metadata Pilot Project con dos objetivos principales:

- mejorar la búsqueda y el acceso en línea a las publicaciones del Australian Commonwealth Government a través de la National Bibliographic Database (NBD) disponible en (<http://www.nla.gov.au/kinetica>) y,
- agilizar la identificación y archivo de estas publicaciones en PANDORA u otro archivo público para asegurar el acceso continuo a ellas a través de NBD.

Un objetivo básico es revisar los estándares de metadatos requeridos y desarrollar y promulgar directrices para animar a las agencias participantes a crear metadatos de calidad.

El Proyecto Pilot involucra a varias agencias gubernamentales; continuará durante el 2005 y se divide en las siguientes fases:

- exploración de opciones (incluyendo la recolección) para la adquisición de metadatos de las agencias colaboradoras
- conversión de los metadatos recogidos del formato Dublin Core al formato MARC21 mediante MODS para su inclusión en NBD.
- automatización de la recolección, conversión, carga y actualización de metadatos; y,
- mejora del software PANDAS para poner en funcionamiento la recogida y archivo automatizados de los recursos web asociados con los metadatos (esperamos que esta fase se desarrolle durante 2005)

(iv) Explorar y promover métodos para archivar las publicaciones web obtenidas por recolección.

Ya está disponible información útil sobre este tema a través de PADI y, tal y como hemos mencionado anteriormente, la conferencia sobre archivo web proporcionará un vehículo para la exploración en profundidad de las actuales prácticas y recientes desarrollos en este tema. Además, se realizará una encuesta sobre las actividades actuales de una serie de agencias culturales para asegurar que están igualmente representadas en PADI.

Durante el 2005 se reestructurará PADI para hacer más fácilmente accesible la información que contiene. La reestructuración se dirigirá a algunos temas actuales del sistema que afectan a la búsqueda de recursos. El objetivo será hacer a PADI más útil mediante la reorganización de los recursos clave de acuerdo a los temas más habituales – incluyendo el Archivo Digital – y haciendo estos temas (que serán denominados Rutas (Trails) accesibles desde la página principal de PADI.

Preservación de materiales digitales (Acción 3.4 de ICABS)

El objetivo de nuestras actividades en éste área de acción es analizar los estándares, directrices y códigos existentes para la preservación de materiales digitales, en cooperación con la Sección de Preservación y Conservación de IFLA.

El desarrollo de programas para la preservación de materiales digitales es un desafío que intimida a muchos bibliotecarios. El propósito de nuestras actividades en éste área es identificar los factores que influyen en la buena disposición de las bibliotecas para involucrarse, e identificar las formulas de asesoramiento que las bibliotecas pueden usar para mejorar su predisposición o para hacer más efectivos los programas que utilizan.

El empleo de PADI para identificar los documentos que puedan servir de guía existentes y hacerlos más accesibles es primordial para esta actuación. También se identificarán importantes carencias y se incluirán en PADI a mediados del 2004. Como se mencionó anteriormente, las estructuras e interfaces de PADI están siendo reestructurados para facilitar información en los temas principales como estándares y directrices, para hacerlos más accesibles.

La relación de la Biblioteca con otras bibliotecas en las regiones del Sudeste Asiático y el Pacífico nos proporcionará una excelente oportunidad para la evaluación de las fórmulas de asesoramiento que pueden ser más útiles en una región amplia y diversa. La Biblioteca funciona como Centro Regional para Oceanía y el Sudeste Asiático para el Programa de Preservación y Conservación de IFLA y ha establecido vínculos con la campaña de la División sobre Sociedad de la Información de UNESCO para la preservación del patrimonio digital.

En Noviembre de 2002, UNESCO acogió una Consulta Regional sobre preservación digital conjuntamente con la preparación de las directrices que la Biblioteca Nacional de Australia se había comprometido a escribir. Basándonos en la información recogida, en Febrero de 2004 preparamos un cuestionario sobre los factores de preparación y unos consejos para mejorarla, para la propuesta del seminario Memoria del Mundo de UNESCO celebrado en Manila, Filipinas. Los factores de preparación incluyen, la existencia de una colección de recursos digitales, el acceso al conocimiento y destrezas

de la Tecnología de la Información; y la dedicación de recursos humanos para la gestión de la colección digital.

Estos documentos se revisarán a la luz de las reacciones recibidas durante el seminario y preparados para su difusión en junio de 2004. Por supuesto, también estarán disponibles en PADI.

Traducción de Elena Escolano
Biblioteca Nacional de España