



## Archivo web internacional: ¿Interoperabilidad en el futuro?

**Grethe Jacobsen**, dr. phil.  
Jefe de Depósito Legal y Mapas,  
Impresos y Fotografías,  
The Royal Library – The National Library of Denmark,  
P.O. Box 2149, DK-2149 Copenhagen K,  
Dinamarca  
[www.kb.dk](http://www.kb.dk)

*Traducción al español: M<sup>a</sup> Jose Llovera Lozán  
Biblioteca Nacional de España*

**Reunión:** 73 Bibliotecas Nacionales  
**Interpretación simultánea:** No

**WORLD LIBRARY AND INFORMATION CONGRESS: 73RD IFLA GENERAL CONFERENCE AND COUNCIL**

19-23 August 2007, Durban, South Africa  
<http://www.ifla.org/iv/ifla73/index.htm>

### Resumen

Algunas bibliotecas nacionales están capturando partes de Internet o están pensando hacerlo, pero para ofrecer una visión completa de Internet los archivos web deben ser interoperables y permitir que el usuario pueda realizar búsquedas integradas. Se ha enviado un cuestionario sobre este tema a 95 bibliotecas nacionales. Las respuestas muestran acuerdo en este objetivo y en que el archivo web es algo cada vez más común. La coparticipación es el elemento clave para avanzar y se ha sugerido una distinción entre los participantes que tienen que ver con la conservación (archivos, museos) y los participantes técnicos (empresas privadas, universidades, otras instituciones de investigación). Trabajar con empresas privadas, con ánimo de lucro, puede forzar también a las bibliotecas nacionales a dejar espacio a maneras de hacer y de pensar nada ortodoxas. El mayor reto en estos momentos es adaptar al entorno de Internet el depósito legal, el copyright y otras legislaciones, para así ser capaces de preservar y dar acceso a las generaciones presentes y futuras.

### Antecedentes y cuestionario

Muchas bibliotecas nacionales han empezado a capturar partes de Internet, ahora consideradas en su mayoría como parte del patrimonio cultural nacional. Sin embargo, dada la naturaleza de Internet, un archivo web nacional que almacene sólo partes de Internet nunca podrá ofrecer una visión completa de Internet, ya que

estuvo disponible y fue utilizada por sus ciudadanos en un momento dado. El objetivo a largo plazo debería ser que todos los archivos web nacionales fuesen interoperables, de manera que cada usuario de cualquier biblioteca nacional (u otra biblioteca) pudiera hacer búsquedas integradas en todos los archivos web.

Poner en práctica este objetivo implica no sólo cuestiones técnicas y de financiación, sino también cuestiones legales como copyright, materiales publicados ilegalmente en Internet y protección de datos personales, especialmente en lo que se refiere a dar acceso a materiales capturados.

Esto requerirá no sólo de una estrecha cooperación entre bibliotecas nacionales sino más bien de colaboraciones en el reparto de los costes para el desarrollo técnico del trabajo de captura, dando acceso, compartiendo las colecciones y preservando los materiales online. Igual de importante es la necesidad de las bibliotecas nacionales de actuar conjuntamente para pedir a los legisladores el permiso para capturar y dar acceso a los archivos, así como en la negociación con los propietarios del copyright.

Como primer paso hacia la consecución de este objetivo, The Royal Library, la Biblioteca Nacional de Dinamarca, envió un cuestionario en marzo de 2007 a las listas de correo de las bibliotecas nacionales del CDNL y CENL, un total de 95 bibliotecas.

Recibimos 23 respuestas antes de la fecha límite (el 27 de marzo) y otras 14 fuera de plazo pero a tiempo de ser incorporadas en este informe (una respuesta del 39%). No todos los que contestaron trataron o pudieron tratar todas las cuestiones así que las respuestas a algunas de las preguntas son pocas. Las respuestas se enriquecieron mucho gracias a los comentarios que permitía incluir el cuestionario y que demuestra que los que contestaron están muy involucrados en el tema y creen en el propósito del archivo web.

A continuación se plantean las preguntas y los comentarios acompañados de una nota referida a la práctica danesa, no tanto para promocionar nuestra propia experiencia como para clarificar lo que esto significa a partir de un ejemplo práctico.

Resumen de la experiencia danesa en general: hemos estado recogiendo información del dominio .dk desde julio del 2005 cuando entró en vigor la nueva ley de Depósito Legal. Nuestro objetivo es conservar la parte danesa de Internet como parte del patrimonio cultural para las generaciones futuras; sin embargo, solos no podemos duplicar en su totalidad la navegación típica actual de Internet a menos que podamos proporcionar acceso a otras partes de la web a otros archivos web.

## **Preguntas y respuestas**

La primera pregunta tenía como título:

### **1. La interoperabilidad en general**

y la pregunta era:

**¿Está de acuerdo con la declaración “El objetivo a largo plazo debería ser un acceso integrado a los materiales de Internet pasados y presentes para todos los ciudadanos más allá de las fronteras nacionales”?**

33 estuvieron de acuerdo con esta propuesta y 4 en contra. Siete hicieron comentarios a la pregunta, sólo una de las respuestas estaba en contra de la propuesta y el resto consideraba que aunque el objetivo valía la pena era poco realista o por lo menos inalcanzable en estos momentos debido a problemas legales y a las diferentes políticas de colecciones. El comentario en contra merece una mención:

“No estoy de acuerdo en que el acceso necesite o deba ser ‘integrado’. La fuerza de Internet está en su heterogeneidad. Creo que un intento de imponer una solución homogénea puede reducir la innovación y dejar de ser benéfica.”

El argumento no es tanto contra el objetivo indicado (el acceso integrado) como contra una forma de cooperación que pudiera limitar las estrategias y las soluciones innovadoras. El primer objetivo de un trabajo de cooperación sería, por tanto, asegurar un desarrollo continuo y que éste dejara espacio para hacer y pensar de una manera no ortodoxa.

Esto es todo un reto para las bibliotecas nacionales, especialmente para sus directores. Tendrán que enfrentarse – en primer lugar – a las peticiones de las autoridades públicas y los políticos para encontrar soluciones al momento y generar resultados para mostrar qué se ha hecho con todo el dinero destinado al archivo web y – por otro lado– de aquellos que trabajan en el archivo web que demandan poder consultar varios sitios al mismo tiempo, probar diferentes métodos y sobre todo dejar espacio a las pruebas de ensayo / error.

La primera pregunta tenía como título:

**2. Archivo web – ¿en la actualidad o en planificación?**

y las preguntas eran las siguientes:

**¿Dispone de un archivo web en la actualidad?**

19 respuestas contestaron Sí a la pregunta sobre si estaban trabajando en la actualidad con los archivos web; 18 contestaron que no, de los cuales 11 afirmaron que iban a empezar a trabajar el archivo web y 7 dijeron que no. Esto significa que el archivo web ya no es una tarea exclusiva de unos pocos, como era el caso hace unos años.

De los que están capturando o están planeando hacerlo, 15 capturan (o piensan hacerlo) el dominio entero, 27 recopilan (o planean hacerlo) páginas web seleccionadas dentro del dominio nacional, 13 también capturan (o piensan hacerlo) fuera de los dominios nacionales y 8 capturan (o piensan hacerlo) páginas web según el idioma.

En Dinamarca, la Ley de Depósito Legal nos permite capturar materiales publicados en el interior del nivel de dominio .dk así como materiales publicados en otros dominios de Internet que están dirigidos a un público en Dinamarca, así que capturamos el dominio entero así como las páginas seleccionadas fuera del dominio. Hemos encontrado unas 30.000 páginas

web fuera del dominio .dk que están dirigidas a una audiencia danesa, principalmente sitios con texto en danés pero también sitios que pertenecen a empresas danesas o a instituciones, o a individuos (p. e. músicos) que viven en Dinamarca.

Entonces, pedimos información sobre las actividades de archivo web, especialmente si los participantes habían creado páginas web con esta información y si era así, les pedíamos que nos diesen la URL. Los títulos y las preguntas eran las siguientes:

### 3. Información de actividades

**3.a.** ¿Dispone de página web con información sobre sus actividades de archivo web?

**3.b.** ¿Dispone de una política sobre la captura de materiales de Internet?

**3.c.** ¿Dispone de una política sobre el descubrimiento y la inclusión de páginas web de relevancia?

15 respuestas contestaron que sí tenían una página web con información sobre sus actividades. Un examen más exhaustivo de estas páginas demostró que la mayoría tienen una versión en inglés, así que claramente el archivo web se ve como una tarea internacional y que la información no va dirigida únicamente a la audiencia del propio país sino también a compañeros de profesión y a otras partes interesadas en el extranjero. 4 tenían página web en su propia lengua y uno tenía una página web en alemán y francés.

Compartir políticas es otro asunto. Todos los que realizan capturas de manera activa tienen una normativa o la están formulando.<sup>1</sup> Sin embargo, sólo 6 de los que respondieron la tienen accesible online. Esto puede deberse más a una falta de tiempo para la traducción que a una decisión de no tenerlas disponibles para el público en general.

Ciertamente, éste es el caso de la Biblioteca Nacional de Dinamarca. Mantenemos una web bilingüe [www.netarkivet.dk](http://www.netarkivet.dk) (danés) y <http://netarkivet.dk/index-en.php> (inglés) y hemos decidido traducir y publicar todas nuestras políticas al inglés además de al danés. Hasta ahora sólo hemos publicado una.

La cuarta pregunta tenía que ver con cuestiones legales para la captura en la legislación (u otra regulación del gobierno) o en los acuerdos voluntarios

### 4. Normas para la captura

¿Realiza capturas... (Elija una o más):

Siguiendo la legislación?  (Especificar p. e. Ley de Depósito Legal):

Con acuerdos voluntarios con otras instituciones?  (Especificar):

Con acuerdos voluntarios con comunidades?  (Especificar):

---

<sup>1</sup> Una respuesta que dispone de archivo web no contestó a esta pregunta

Con acuerdos voluntarios  
con organizaciones  (Especificar):  
privadas o personas?

Otros?  (Especificar):

11 respuestas capturaban siguiendo una ley de depósito legal, 1 siguiendo otras normativas legales, 1 cumpliendo la ley de copyright y otro una ley de datos personales como algo excepcional. Además, una de las respuestas capturaba el dominio nacional como parte de un proyecto. 4 respuestas capturaban siguiendo acuerdos con editores, otras comunidades y el Internet Archive y 1 optó por no hacerlo.

Parece que todavía existe una manera de dirigirse a esos países que quieren realizar sus capturas utilizando las leyes de Depósito Legal u otro tipo de legislación, pero más de la mitad de los que afirmaron estar trabajando en archivo web lo hacen bajo legislación de Depósito Legal mientras que tres lo hacen mediante otro tipo de legislación. Cabe señalar que hay una biblioteca activa en temas de archivo web, la Biblioteca Nacional de Holanda, que no tiene Depósito Legal pero que sin embargo tiene una larga tradición de negociación con editores para depositar el patrimonio nacional tanto online como en formato físico.

La ley de Depósito Legal de Dinamarca que permite la captura de partes danesas de Internet se aprobó en diciembre de 2004 y entró en vigor en julio de 2005. Anteriormente, podíamos capturar publicaciones de la web desde 1998 hasta 2005 siguiendo la anterior ley de Depósito Legal (en vigor desde 1998 hasta 2005). Además, teníamos permiso del Ministerio de Cultura para realizar diferentes tipos de captura desde 2001 a 2004 como parte de varios proyectos mientras se preparaba la nueva ley de Depósito Legal.

Una cuestión importante relativa al archivo web será su integridad y, por tanto, preguntamos si las bibliotecas que trabajaban en archivo web podían deshacerse de materiales capturados y si esto podía controlarse.

## 5. Manipulación del contenido del archivo

5.a. ¿Qué cantidad de los materiales capturados almacena realmente en su archivo web?

Todo lo capturado

Sólo parte de lo capturado  (Especificar):

5.b. Si se eliminan de materiales capturados, ¿puede controlarse?:

Sí  (Especificar):

No

21 respuestas afirmaron guardar todo lo capturado y sólo 1 respuesta manifestó que su biblioteca sólo guardaba parte de lo capturado, archivos de menos de un cierto límite. No se especificaba el límite en la respuesta, ni en la web de la biblioteca. En general, parece que las bibliotecas tienden a mantener sus archivos intactos.

El acceso es el elemento básico de un archivo web – o en cualquier archivo para este propósito. Si no hay acceso al archivo, realmente no hay razón por la que capturar ya que el propósito de un archivo web es documentar parte del patrimonio cultural de una nación. Además, el acceso debería garantizarse a todo el mundo como el acceso a todos los demás materiales pertenecientes al patrimonio cultural de una nación.

En este sentido, preguntamos lo siguiente:

#### **6. Acceso a los materiales de archivos**

¿Permite un acceso online generalizado a su archivo web?

*Si su respuesta es “no”, conteste a la siguiente pregunta:*

Acceso limitado:	(Por favor, especifique. P. e. sólo investigación, objetivos estadísticos)
Sin acceso:	

8 contestaron sí a la pregunta “¿Permite un acceso online generalizado a su archivo web?” Sin embargo, 3 de ellos apuntaron que, técnicamente, el acceso al público todavía no era posible. Entre el resto había varios tipos de acceso en línea y fuera de línea, incluyendo el acceso con permiso de los editores (6), acceso con permiso (4) y acceso sólo para investigadores (4). Podemos concluir que el acceso online generalizado a los archivos web todavía no es muy común por motivos legales (copyright y protección de datos) y por motivos técnicos.

En Dinamarca, sólo permitimos el acceso por motivos de investigación o estadística y por tanto sólo a investigadores y a niveles de post doctorado. La razón no es, como se pudiera pensar, la legislación por copyright sino por el Acta Danesa de Proceso de Datos Personales. Vamos a iniciar conversaciones con la Agencia para permitir el acceso para al menos algunas partes de los datos recogidos, por ejemplo páginas web de instituciones públicas. Nuestro objetivo principal era permitir un acceso generalizado en las Salas de Lectura de la *Royal Library* y la *State and University Library*, para que esta parte del patrimonio cultural fuese accesible a todos los ciudadanos, como a todos los libros y otros tipos de materiales publicados tiene acceso todo el mundo, ya sea bajo préstamo o utilizados con la normativa de la biblioteca.

El intercambio de materiales capturados, de URLs que se recogen u otra información de cualquier archivo web puede ser de interés para otros archivos y, por lo tanto, un elemento importante en la cooperación internacional de los archivos web. La pregunta 7 trataba este asunto:

#### **7. Captura ilimitada de materiales**

¿Puede proporcionar copias de los materiales capturados a otros archivos web nacionales?

*Si la respuesta es “Sí”, conteste a la siguiente cuestión:*

¿Puede proporcionar (por favor, marque la casilla adecuada):

- |   |                          |
|---|--------------------------|
| Copia de materiales capturados                                | <input type="checkbox"/> |
| Información sobre la recogida de URLs y momento de la captura | <input type="checkbox"/> |
| Otro tipo de información                                      | <input type="checkbox"/> |

Sólo 3 de las respuestas podían ofrecer copias de los materiales capturados, 2 no lo sabían y 15 respondieron que no. 8 respuestas podían ofrecer copias de las URLs recogidas y dar un tiempo de captura de los materiales y 1 no tenía claro este tipo de cooperación internacional. Además, dos respuestas apuntaban que los metadatos y los registros catalográficos para algunos de los materiales capturados pueden encontrarse en su catálogo online.

En Dinamarca, sólo podemos ofrecer copias e información sobre lo que hemos capturado a los investigadores que han obtenido un permiso para poder acceder al archivo. No tenemos ningún registro en nuestro OPAC ni ningún plan para tener parte o todo el archivo web catalogado como registros en el OPAC. Sí que tenemos registros para las publicaciones web de lo que capturamos entre 1998 y 2005 pero no podemos ofrecer acceso a esas colecciones en estos momentos, al formar parte del archivo web y estar, por tanto, protegidas por el Acta Danesa de Proceso de Datos Personales.

Otro aspecto de la cooperación internacional es la ayuda entre las partes para la búsqueda de materiales que puedan formar parte del patrimonio cultural nacional, difíciles de encontrar sin ayuda de los colaboradores. Por tanto, la pregunta que hicimos fue si este tipo de colaboración era posible.

#### **8. Cooperación con otros archivos web**

¿Podría (marcar la casilla apropiada):

- |   |                          |
|---|--------------------------|
| Ofrecer información sobre las URLs de su dominio nacional que son de interés para otras bibliotecas nacionales? | <input type="checkbox"/> |
| Capturar páginas web de su dominio nacional que son de interés para otras bibliotecas nacionales?               | <input type="checkbox"/> |
| Permitir que otras bibliotecas nacionales y sus usuarios accedieran a su archivo?                               | <input type="checkbox"/> |

14 respondieron que podían ofrecer información sobre las URLs de su dominio nacional que pudieran ser de interés para otras bibliotecas nacionales, 13 también podían capturar páginas web y 10 permitían el acceso a otras bibliotecas nacionales.

El archivo web danés no puede ofrecer información sobre URLs si se refiere a materiales capturados, ni recopilar páginas web para otras bibliotecas

nacionales. Sí podemos, por supuesto, ofrecer información sobre URLs y contenido que es de disponibilidad pública en el dominio .dk si nos hemos encontrado con esa información.

Muy relacionado con estos temas de cooperación está la elección de los colaboradores que era el enunciado de la novena y última pregunta, bajo el título:

## 9. Colaboraciones en la captura de materiales de Internet

y planteamos las siguientes preguntas:

**9.a.** ¿Realiza la captura en colaboración con otras instituciones?

*Si la respuesta es "Sí", conteste a la siguiente cuestión:*

**9.b.** ¿Trabaja con (marcar la casilla apropiada):

Otras bibliotecas nacionales	<input type="checkbox"/>	
Otras bibliotecas	<input type="checkbox"/>	
Otras instituciones públicas	<input type="checkbox"/>	Especificar el tipo (archivo, museo, universidad, etc.)
Comunidades	<input type="checkbox"/>	Especificar
Empresas privadas / instituciones	<input type="checkbox"/>	Especificar el sector (IT, comercio, publicidad, ect.)
Individuos	<input type="checkbox"/>	

13 de los que realizaban capturas contestaron que sí a esta cuestión y también ofrecieron más detalles de esta colaboración.

4 colaboraron con otras bibliotecas nacionales, 9 con otras bibliotecas, 8 con otras instituciones públicas, 4 con comunidades y 4 con empresas privadas o instituciones. Como indican las cifras, muchas de ellos trabajaron con más de un tipo de institución; la colaboración ha recurrido con más frecuencia a las bibliotecas, otras instituciones que se mencionan son el Internet Archive, universidades, archivos nacionales, archivos audiovisuales, museos, compañías de publicidad, una fundación privada. Por los comentarios se deduce que lo que marca la dirección de las bibliotecas en la elección de sus colaboradores es el compromiso de esos colaboradores con el archivo web y/o su pericia técnica.

La colaboración también trae consigo complicaciones en términos de quién es el propietario del archivo y a la pregunta:

**9.c.** ¿Cómo ha resuelto el problemas de propiedad del archivo y sus contenidos?

4 respondieron que habían compartido la propiedad, 7 afirmaron que su institución era la propietaria del material y 1 respondió que no lo habían determinado del todo, 3 tenían diversos acuerdos con sus colaboradores.

Entre los costes de construir y mantener un archivo web están los costes de desarrollo, que también es un factor importante en las bibliotecas que hayan pensado en la colaboración.



A la pregunta

**9.d.** ¿Cómo ha resuelto los problemas técnicos?

Desarrollo conjunto  
Mi institución es responsable  
Mi colaborador es el responsable  
Otro tipo de solución

	Especificar

6 contestaron que estaban comprometidos en un proyecto conjunto con colaboradores, 4 instituciones están a cargo del desarrollo, 2 lo dejan en manos del colaborador, 1 se apoya en un software desarrollado por IIPC (International Internet Preservation Consortium).

En Dinamarca, las dos bibliotecas comprometidas con el archivo web, *The Royal Library* (la Biblioteca Nacional) y la *State and University Library*, han creado una institución virtual, "netarkivet.dk" (netarchive.dk) que está dirigida por un Comité Directivo de 6 miembros (3 de cada biblioteca) expertos en tecnología web. El Comité se reúne unas dos o tres veces al año para discutir y decidir sobre políticas y cuestiones legales sobre aspectos económicos y técnicos. Netarkivet.dk tiene un gestor dependiente del Comité Directivo y que supervisa el trabajo diario. El desarrollo es compartido para las dos bibliotecas, en parte con los colaboradores en el IIPC.

También preguntamos lo siguiente a aquellos que no cooperaban con otras instituciones en la captura:

Si tuviera que colaborar para la captura de materiales de Internet, ¿con qué tipo de institución le gustaría trabajar y qué solución adaptaría en temas de propiedad del archivo, los contenidos y las cuestiones técnicas?

En la actualidad, muchos de los que están trabajando con archivos web, también respondieron a esto. Los comentarios dejaron entrever la necesidad de colaborar con instituciones que tengan experiencia tanto técnica como en colecciones para ofrecerla junto a un compromiso con la preservación. En algunos casos, las bibliotecas distinguían entre colaboraciones técnicas y colaboraciones de conservación. Las colaboraciones técnicas mencionadas provienen de universidades, instituciones de investigación y empresas privadas y colaboradores de conservación: archivos, museos y otras fundaciones de preservación. Como dice la Biblioteca Nacional de Australia: "contra más trabajo conjunto pueda hacerse, mejor".

En Dinamarca, la palabra clave que apareció desde que se habló por primera vez de archivo web fue colaboración. Estamos activos en el IIPC y en iniciativas europeas y proyectos y también trabajamos muy de cerca con otros países nórdicos ya que todos trabajan con archivos web.

### **Conclusiones:**

Todas las respuestas apoyan la idea del archivo web aunque para algunos no parezca realista que llegaremos a poder almacenar Internet al completo. El reto para las bibliotecas nacionales es hacer que este objetivo sea alcanzable.

Las colaboraciones y la cooperación es un elemento clave para avanzar. Esto no es algo nuevo para las bibliotecas, lo nuevo es que los colaboradores provienen de diversos ámbitos tanto públicos como privados. Parece útil la distinción entre colaboradores técnicos y colaboradores de conservación. Para encontrar colaboradores de conservación sería necesario buscar más allá de las bibliotecas, en archivos, museos y otras instituciones con funciones de depósito. Probablemente todo sea público. Los colaboradores técnicos podrían ser los mismos que los de conservación, pero lo que más abunda son instituciones privadas, universidades y otras instituciones de investigación. Trabajar con empresas privadas, con ánimo de lucro, puede tener la ventaja añadida que las bibliotecas nacionales se verán obligadas a mentalizarse de que el archivo web debe ir acompañado por una investigación y un desarrollo continuo que también debe dejar espacio a maneras de hacer y de pensar no ortodoxas. El reto para las bibliotecas nacionales es encontrar un colaborador para cada función a desarrollar. Sin embargo, el mayor reto para las bibliotecas nacionales ahora mismo es conseguir que las leyes de Depósito Legal, el copyright y otras legislaciones se adapten a Internet, para poder preservar esta parte del patrimonio nacional y ponerlo a disposición de las generaciones presentes y futuras.

Al final del cuestionario, pedimos comentarios sobre el cuestionario y sobre las preguntas del mismo y una de ellas nos ofrece una conclusión perfecta a este informe y una exhortación a IFLA y a las bibliotecas nacionales: "Esperamos que este cuestionario nos ayude a todos de manera que desde IFLA se difundan los problemas y las cuestiones legales relacionadas con el archivo web."