

ARTÍCULO

Preservación de objetos de aprendizaje en repositorios digitales

Juanjo Boté

juanjo.botev@ub.edu

Departamento de Biblioteconomía y Documentación de la Universidad de Barcelona

Julià Minguillón

jminguillona@uoc.edu

Universitat Oberta de Catalunya

Fecha de presentación: noviembre de 2010

Fecha de aceptación: julio de 2011

Fecha de publicación: enero de 2012

Cita recomendada

BOTÉ, Juanjo; MINGUILLÓN, Julià (2012). «Preservación de objetos de aprendizaje en repositorios digitales» [artículo en línea]. *Revista de Universidad y Sociedad del Conocimiento (RUSC)*. Vol. 9, n.º 1, págs. 22-35 UOC. [Fecha de consulta: dd/mm/aa].

<<http://rusc.uoc.edu/ojs/index.php/rusc/article/view/v9n1-bote-minguillon/v9n1-bote-minguillon>>

ISSN 1698-580X

Resumen

El propósito de este artículo es analizar los diferentes procesos en la conservación de objetos de aprendizaje en un repositorio digital. Como caso de estudio se presenta un prototipo de repositorio basado en una colección de materiales de *e-learning* sobre estadística, usados en las asignaturas respectivas de la Universitat Oberta de Catalunya. Estos materiales tienen el propósito de servir a toda la comunidad, no tan solo a la universidad. Para ello, se ha creado este repositorio en una plataforma abierta basada en DSpace. El propósito es promover tanto la reutilización como la conservación digital de dichos materiales de *e-learning*, aunque ambos objetivos son, en ciertos aspectos, contradictorios. En este artículo se analizan los requerimientos de los objetos de aprendizaje depositados en un repositorio y las necesidades de los diferentes roles que intervienen en su manipulación y su conservación a largo plazo.

Palabras clave

objetos de aprendizaje, repositorios, preservación digital, metadatos, archivos, bibliotecas digitales

Preservation of Learning Objects in Digital Repositories

Abstract

The aim of this article is to analyse the different processes involved in the preservation of learning objects in a digital repository. Presented as a case study is a prototype repository for a collection of statistics-related e-learning materials used in the respective academic subjects offered by the Open University of Catalonia (UOC, Universitat Oberta de Catalunya). The purpose of these materials is to serve the whole community and not just the university. To that end, the repository was created with DSpace open-source software. The goal is to promote the reuse and the digital preservation of such e-learning materials, though certain aspects of these two objectives are somewhat contradictory. This article analyses the requirements of learning objects deposited in a repository and the needs of the various roles intervening in their handling and long-term preservation.

Keywords

learning objects, repositories, digital preservation, metadata, archives, digital libraries

1. Introducción

En muchas instituciones de educación superior hay una tendencia creciente a emplear entornos virtuales de aprendizaje (EVA). En un EVA, todos los aspectos de un curso están gestionados a través de una interfaz de usuario consistente, que normalmente es estándar en toda la institución. Uno de los elementos habituales de un EVA es un repositorio de objetos de aprendizaje, utilizado para gestionar los recursos docentes utilizados durante el curso. No existe una definición común sobre el concepto de repositorio (Conway, 2008); aun así estos son empleados de forma abierta para facilitar materiales o información a una comunidad determinada. Habitualmente, un repositorio institucional (Shreeves y Cragin, 2008) incluiría, entre otros, informes, publicaciones, cursos completos y manuales, pero también objetos de aprendizaje y datos de proyectos de investigación, etiquetados de acuerdo con algún esquema de metadatos, preferentemente IEEE LOM o Dublin Core (Neven y Duval, 2002).

La preservación digital es un elemento clave del diseño de un repositorio, dado que estos recursos se crean con una versión de software concreta y necesitan actualizarse para asegurar su acceso posterior durante el tiempo que sea necesario, ya que en caso contrario se podrían producir pérdidas de información.

Este hecho conlleva, además, la necesidad de emplear un análisis de riesgos para poder determinar las prioridades de operaciones de preservación digital que se ejecutarán sobre los recursos. El análisis de riesgos también podrá servir para evaluar qué coste es asumible para poder preservar digitalmente versiones de software o materiales de aprendizaje caducados.

Por otra parte, las particularidades de los objetos de aprendizaje utilizados en un EVA hacen necesario replantearse los mecanismos habituales de preservación relacionados con el etiquetado mediante metadatos, dada la gran variabilidad de elementos que se pueden encontrar en el proceso de aprendizaje, tanto por tipología (ejercicios, ejemplos, simulaciones, etc.) como por formato.

Este artículo está organizado de la manera siguiente. La sección 2 describe el repositorio piloto usado como caso de estudio para analizar una solución de preservación digital basada en DSpace. La sección 3 describe los elementos que han de tenerse en cuenta para establecer los criterios necesarios para asegurar la preservación de los objetos depositados. En la sección 4 se presentan las políticas de preservación utilizadas en el repositorio. Finalmente, las conclusiones de este trabajo y las líneas de trabajo actuales y futuras se pueden encontrar en la sección 5.

2. Repositorios digitales abiertos

La Universitat Oberta de Catalunya ha creado un repositorio digital abierto¹ sobre Estadística, siguiendo un proceso de diseño centrado en el usuario (Ferran *et al.*, 2009). Son materiales de aprendizaje en forma de ejercicios, materiales de estudio, documentos multimedia y ficheros con datos de programas de estadística concretos.

Estos recursos están en un gran abanico de formatos, incluyendo Minitab, Word y Portable Document File (PDF), aunque cada vez más existen materiales docentes en formato de vídeo y otros documentos (tanto textuales como de datos), todos ellos en una amplia gama de versiones de software y de sistemas operativos. Debido a esta gran variedad, ha sido necesario no solo pensar en el repositorio para gestionar los materiales de aprendizaje, sino también en la preservación digital a largo plazo (Smith, 2005).

Desde su inicio, la UOC ha generado un considerable número de recursos docentes en estadística. Dichos recursos han sido creados a través de un proceso de publicación de alta calidad, involucrando diferentes roles (autor, editor, profesor, gestor, etc.), y han sido publicados siguiendo un modelo editorial de producción y diseño basado en módulos, los cuales son unidades de aprendizaje de pocos créditos diseñados para el aprendizaje de un tema específico del contenido de un curso. El proceso de creación y edición es de alrededor de un año. El mantenimiento y la actualización de los cambios correspondientes es un proceso costoso y complejo, ya que significa realizar de nuevo todo el proceso editorial o parte del mismo. También han de gestionarse del mismo modo otras tipologías de recursos asociados con las materias en cuestión y que se han generado como resultado de las actividades de enseñanza de cada semestre académico, como ejemplos, ejercicios, experimentos y ejercicios de autoaprendizaje.

En este repositorio se han aplicado los conceptos propios de una biblioteca digital y servicios de referencia a un entorno de aprendizaje virtual. Cuestiones como la búsqueda y los servicios bibliográficos son aspectos similares, pero no idénticos. Los estudiantes están habituados a buscar información en la biblioteca digital por autor o título, pero en el caso de los objetos de aprendizaje hay un número de títulos duplicados de ejercicios, ficheros y unidades de aprendizaje que han obligado a utilizar una taxonomía diferente para clasificarlos, ya que el concepto de título es ambiguo.

El repositorio permite que los estudiantes y otras personas de la comunidad de aprendizaje tengan acceso a los materiales educativos de forma organizada. Estos materiales estaban dispersos entre

1. <http://oer.uoc.edu> (actualmente en desarrollo).

diferentes asignaturas, eran inaccesibles a la comunidad de forma conjunta y no estaban clasificados ni ordenados. En consecuencia, muchos estudiantes se encontraban perdidos en el uso de estos recursos y el profesorado no obtenía el beneficio esperado (Barker *et al.*, 2004). Además, no se disponía de criterios de búsqueda adecuados para identificar aquellos recursos que podrían reutilizarse para crear nuevos materiales educativos. Ejercicios, notas, módulos, artículos, libros de texto, herramientas de software, laboratorios virtuales, recursos de audio, vídeos, planes de estudio, horarios, calendarios, actividades, simulaciones y objetos de aprendizaje, entre otros documentos, son más accesibles si se utiliza un repositorio digital.

2.1. La plataforma DSpace

Hay muchas razones para escoger DSpace² como repositorio. Es una plataforma de código abierto, acepta diferentes esquemas de metadatos, incorpora políticas de conservación a largo plazo, utiliza *handles* para identificar cada elemento de forma perpetua y es sólido y estable. Además, no menos importante es que ya se encuentra en uso en otras partes de la UOC. Como plataforma de código abierto, DSpace tiene una gran comunidad activa de usuarios y desarrolladores, incluyendo instituciones de educación superior y bibliotecas digitales. DSpace también es válido para la preservación digital a largo plazo, ya que acepta políticas para ello. Incluye herramientas como Checksum Checker, que permite verificar la integridad de los flujos de bits (*bitstreams*) de las áreas de almacenamiento; TechMDExtractor, una herramienta para validar los formatos de flujos de bits almacenados y para extraer metadatos de los flujos de bits. También incluye un paso de flujo de trabajo en la preingesta y un flujo de trabajo opcional que valida el formato de cada flujo de bits después de haber realizado la ingesta, facilitando al administrador los metadatos de los ficheros que son inválidos o están mal formados.

DSpace tiene comunidades y subcomunidades (definidas de forma jerárquica), colecciones, ítems, secuencias de flujo de bits y formatos de secuencias de flujo de bits. Las comunidades y subcomunidades constituyen la organización de más alto nivel, mientras que una colección es un conjunto de ítems, como por ejemplo recursos sobre estadística. Cada colección tiene su propio flujo de trabajo, que puede ser definido por la unidad de gestión. Un ítem es una agrupación de un conjunto útil de contenidos y metadatos, añadidos durante el proceso de ingesta o posteriormente. Por lo que respecta al almacenamiento, un flujo de bits es una secuencia de bits que corresponde al contenido de los ficheros de metadatos, mientras que un formato de flujo de bits implica la captura de formatos específicos de ficheros de los cuales se ha realizado la ingesta, siendo posible mejorarlos con herramientas de extracción de metadatos como DROID³ o la herramienta NLNZ Metadata Extraction⁴.

DSpace emplea por defecto metadatos Dublin Core para archivar colecciones. Sin embargo, en el proceso de ingesta es posible añadir otros metadatos usados para preservar el objeto a largo plazo. Una descripción ampliada permite una mejor recuperación de los objetos del repositorio y también

2. <http://www.dspace.org>

3. <http://droid.sourceforge.net>

4. <http://meta-extractor.sourceforge.net>

mejora sus propiedades significativas, identificando cómo se creó el objeto y otros datos similares. Por otra parte, el análisis del uso de los objetos depositados en el repositorio también permite mejorar los metadatos de forma continua (Ferran *et al.*, 2007), detectando y corrigiendo problemas de etiquetado.

3. Preservación digital

La cuestión clave es si el acceso a estos materiales será posible en un periodo específico de tiempo, por ejemplo 20 años (la UOC se creó hace solo 15 años y ya se han encontrado problemas de preservación). La obsolescencia tecnológica, en este caso ficheros de programas de estadística propietarios, tiene que monitorizarse para prevenir cualquier pérdida de información minimizando todos los riesgos posibles. Esto significa que la conservación digital debe plantearse para que la información pueda ser accesible a lo largo del tiempo. La conservación de estos objetos digitales a largo plazo implica un análisis de riesgos para el establecimiento de prioridades y un plan de preservación para facilitar su acceso. Estos elementos permitirían asegurar la reutilización de los materiales de aprendizaje, entre otros.

Habitualmente, en la UOC se producen procesos de ingesta en el repositorio en los meses de enero y julio, después (y antes) de cada semestre académico. En estos momentos del año se debe analizar el repositorio tecnológicamente, para evaluar la necesidad de migrar los objetos digitales introducidos hacia versiones de software más actualizadas.

Antes de decidir qué tipos de materiales se guardan es necesario llevar a cabo un análisis de riesgos para poder determinar los riesgos de obsolescencia, el establecimiento de prioridades en el plan de preservación, así como los costes que la institución puede asumir con la preservación digital.

Mediante el análisis de riesgos, se evaluarán los materiales cuya reedición será menos costosa y aquellos cuya preservación resulte más económica. Este suceso consiste especialmente en el versionado de software que soporta determinados materiales. En nuestro estudio basamos el análisis de riesgos en la obsolescencia tecnológica.

Uno de los instrumentos que se disponen para el análisis de riesgos en repositorios es DRAMBORA (DCC, DPE, 2007), que permite cuantificar y analizar los riesgos de los materiales de un repositorio.

La metodología DRAMBORA basada en el estándar AS/NZ 4360:2004 permite analizar tanto el entorno como los recursos digitales. DRAMBORA es aplicable a colecciones digitales que están o van a estar en repositorios (McLeod, 2008). En la evaluación de los recursos digitales, es posible decidir la prioridad de la conservación digital de los recursos.

DRAMBORA es una metodología aplicable a repositorios abiertos que se compone de seis etapas: identificación del contexto de la organización, documentación del marco regulador, identificación de los activos, actividades del repositorio y sus propietarios, identificación de riesgos, evaluación de riesgos y gestión de riesgos.

Asimismo, al final del proceso es posible crear un gráfico radial donde se puede comparar la media de otros repositorios similares junto con nuestro repositorio de estadística.

Una de las principales cuestiones es cómo van a guardarse los materiales creados digitalmente por diferentes tipos de instituciones. Las instituciones deben guardar su información digital debido a una gran variedad de razones: administrativas, legales, históricas, personales o de valor científico (artículos académicos, tesis electrónicas, disertaciones, etc.) y, obviamente, docentes. Estas instituciones incluyen universidades, centros educativos, bibliotecas, museos, centros de investigación o personas que desean disponer de su colección privada de materiales. Esto implica una gestión especial de la tecnología y la capacidad de acceder a documentos electrónicos antiguos con tecnología nueva (Lee *et al.*, 2002). Es aquí donde aparece la preservación digital, que permite disponer de datos electrónicos accesibles y útiles durante un largo período de tiempo. Los datos electrónicos deben preservar sus propiedades significativas a través del tiempo y ser accesibles a una comunidad de usuarios designada. En lo que respecta a la preservación a largo plazo, debe entenderse que los materiales electrónicos serán accesibles en un futuro, manteniendo todas sus propiedades significativas desde que fueron creados.

3.1. Preservación en un repositorio digital

Tal y como se ha mencionado, tienen que aplicarse técnicas de preservación digital, de forma que todos los materiales de aprendizaje sean accesibles a los estudiantes y a los profesores a través del tiempo. Las estrategias de preservación más comunes son la migración y la emulación. Algunos ejemplos son VERS Encapsulated Object (VEO) (Waugh *et al.*, 2000) y Universal Virtual Computer (UVC) (van der Hoeven *et al.*, 2005), respectivamente.

La migración consiste en la transformación de un objeto electrónico hacia una versión actualizada de su propio formato de fichero, para que sea más sencillo manipular y acceder a la información. Durante el proceso de migración, algunas propiedades significativas pueden perderse (en particular debido a conversiones de software), así que las descripciones de información en el proceso de ingesta deben ser precisas. El principal objetivo de las estrategias de migración es mantener todas las propiedades significativas de un objeto digital desde su creación. Un ejemplo sería la migración de hojas de cálculo de Lotus Notes a Microsoft Excel o mejor aún hacia un fichero estándar XML.

La emulación es el proceso de crear un entorno donde el software obsoleto debe migrar para poder funcionar en una nueva plataforma. El nuevo software debe migrar posteriormente cuando el simulador esté tecnológicamente obsoleto. Son bien conocidos los emuladores para consolas de juego antiguas como Atari o Sony PlayStation. Esto significa preservar la apariencia de la plataforma y su funcionalidad y tener una copia actualizada del objeto original.

Finalmente, otra aproximación para la preservación a largo plazo es el modelo Open Archival Information System (OAIS), adoptado como ISO 14721:2003 (CCSDS, 2002). Su principal objetivo es la preservación de información para una comunidad designada por un periodo de tiempo indefinido. El modelo OAIS es hoy en día un estándar de referencia en los sistemas de archivos.

3.2. El modelo de referencia OAIS

El modelo OAIS está aceptado ampliamente por instituciones como un modelo de archivo tanto digital como en otros soportes. El modelo OAIS se define como: «Un archivo, consistente en una

organización de gente y sistemas que han aceptado la responsabilidad de preservar la información y ponerla a disposición de una comunidad designada».

Se basa en procesos de información que facilitan una descripción a alto nivel de la información de los objetos gestionados por el archivo o repositorio. Los componentes funcionales de un archivo digital son la ingesta, el almacenamiento de la información, la planificación de la preservación, el acceso, la gestión de los datos y la administración. El modelo OAIS tiene tres actores implicados: el productor, el gestor y el consumidor.

El productor se define como un conjunto de gente o sistemas que facilitan información para que sea preservada. Esto puede incluir otros OAIS, o personas internas del OAIS o sistemas.

Originalmente, el consumidor pertenece a la comunidad designada, que interactúa con los servicios de OAIS para encontrar información preservada de interés y acceso a esta información en detalle. Esto puede incluir otros OAIS, así como personas internas del OAIS o sistemas. Una clase de consumidor de la comunidad designada debería ser capaz de entender la información preservada.

El gestor es un rol que desempeñan los que implementan las políticas globales sobre el OAIS como un amplio componente de las políticas de dominio. En otras palabras, el control de la gestión de OAIS solo es una de las responsabilidades de gestión.

La gestión no está implicada en el día a día de las operaciones del archivo y puede asignar políticas en el repositorio, como por ejemplo el cambio de rol tanto del productor como del consumidor. La figura 1 muestra el modelo OAIS, tal como definimos el modelo clásico de una biblioteca digital. En este caso, los productores crean el contenido y, a través del flujo de trabajo de gestión, el consumidor puede recuperar los temas desarrollados por el productor.

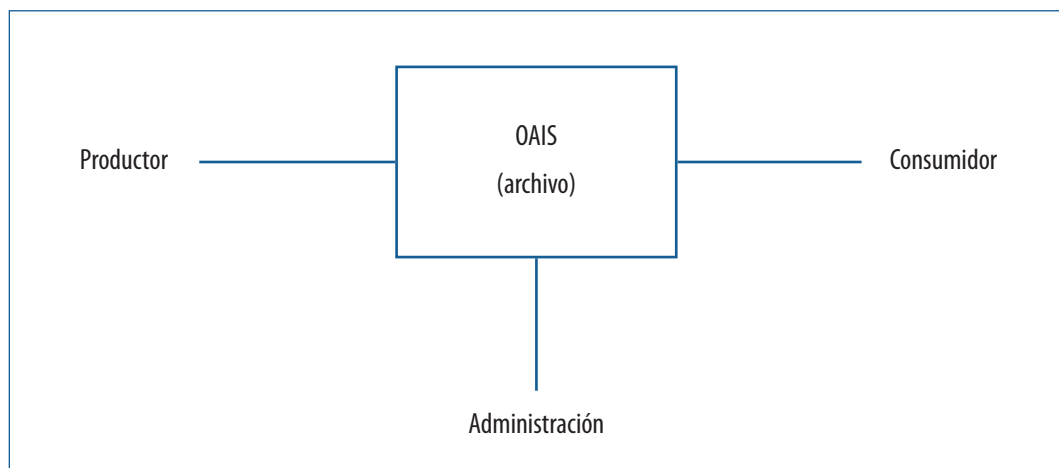


Figura 1. Modelo OAIS (reimpreso con permiso del Consultative Committee for Space Data Systems).

La información se guarda y gestiona en el archivo OAIS. Los consumidores y los productores de la comunidad designada interactúan con los servicios de OAIS para encontrar y adquirir información preservada de interés.

El modelo OAIS permite que los flujos de trabajo se sigan en un archivo digital nuevo. Su implementación significa que se necesitan unos estándares claros para la preservación digital, así como la

creación de un conjunto de terminologías que puedan ser comprendidas por la comunidad designada y la clarificación de procedimientos para disponer de un servicio de archivo fiable.

Los componentes funcionales de un archivo digital son la ingesta, el almacenamiento de información, la planificación de la preservación, el acceso, la gestión de los datos y la administración.

La ingesta se define como un servicio abierto de archivo. Es el proceso de aceptación de un objeto digital por los productores y consumidores seleccionados de acuerdo con el concepto de aprendizaje colaborativo e introduciéndolo en un archivo digital. En nuestro modelo, la función de ingesta incluye la recepción de paquetes de información de *submission* (SIP) por los productores y consumidores seleccionados y la preparación de contenidos para almacenar y gestionar junto con el archivo

El almacén de archivo facilita servicios y funciones de los paquetes de información de archivo (AIP).

La planificación de la preservación facilita los servicios y funciones para asegurar que la información almacenada en el OAIS permanece accesible para la comunidad designada a largo plazo, incluso si el entorno de computación está obsoleto.

Estos actores están implicados en algunos de los componentes funcionales. En un EVA puede disponerse de una visión distinta al modelo OAIS, en el cual los tres actores tienen un papel unívoco. En cambio, en un EVA todos los roles pueden intercambiarse. Profesores y estudiantes pueden ser a la vez productores y consumidores. Los profesores pueden ser productores, consumidores y gestores. A modo de ejemplo, un estudiante puede solucionar un problema de estadística y el profesor puede considerar apropiada su solución para incorporarla al repositorio, adoptando el rol de gestor. Esta variante dentro del modelo OAIS puede definirse a través de la gestión de los flujos de trabajo (Chen, 2004) que soporta DSpace.

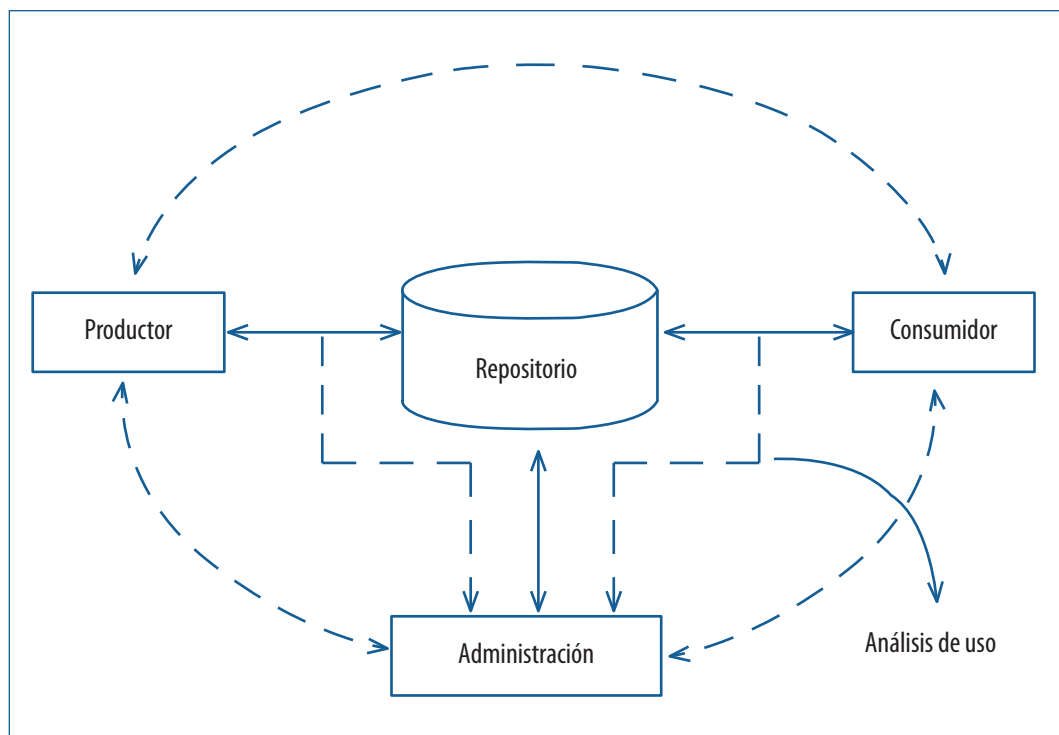


Figura 2. Modelo OAIS en EVA

La figura 2 muestra un posible escenario para el modelo OAI en un entorno virtual de aprendizaje, donde los productores pueden tener el rol de consumidores, y los consumidores pueden tener el rol de productores. Debido a la posibilidad de configuración del flujo de trabajo en el repositorio, es posible asignar derechos de gestión tanto a productores como a consumidores.

Si aplicamos el concepto de web 2.0 a nuestro OER, los conceptos como aprendizaje colaborativo y participación de la comunidad en un modelo OAI clásico pueden cambiar, tal y como se ha visto en la figura 1. Es importante observar también que se mejora el análisis de uso (Lee, 2002). El análisis de uso serviría para cambiar roles entre productor y consumidor dependiendo del uso de sus recursos.

4. Plan de preservación para materiales docentes

Antes de preservar un objeto de aprendizaje deben conocerse sus propiedades significativas, en particular aquellas que hacen que el objeto sea reutilizable, como por ejemplo el formato de fichero. La colección de recursos de estadística que opta a formar parte del repositorio incluye vídeos MPEG, ficheros Microsoft Excel, ficheros Minitab, ficheros SPSS, documentos de texto, documentos LaTeX, ficheros Microsoft Word en diferentes versiones, ficheros PDF, presentaciones en Microsoft PowerPoint y, más recientemente, ficheros en Open Office en versiones muy variadas. Por ejemplo, uno de los retos principales son los ficheros en código propietario tales como SPSS o los ficheros Minitab con fórmulas asociadas. Es posible exportar ficheros Minitab hacia XML o ficheros etiquetados, mientras que SPSS es el estándar *de facto* en los formatos de preservación.

4.1. Propiedades significativas

Las propiedades significativas son una característica de los objetos digitales que deben mantenerse para que sean accesibles a largo plazo (Ashley, Davis y Pinsent, 2008). Las propiedades significativas pueden clasificarse en contenido, contexto, apariencia, estructura y comportamiento. Por razones de espacio este artículo solo examina los aspectos de la primera categoría. Se necesita una descripción de los ficheros mediante metadatos para preservarlos. Estos metadatos mejorarán la descripción de los objetos de los que se ha realizado la ingesta en el repositorio. Los metadatos deberán someterse a un control de calidad para evitar ruido en el proceso de recuperación.

Antes de la ingesta, deberían comprobarse todos los materiales para poder garantizar su preservación y, al mismo tiempo, su reusabilidad tecnológica. Así, por ejemplo, las etiquetas de un fichero PDF/A deben comprobarse cuidadosamente para validar que el fichero tiene un formato correcto y sea coherente. Es aconsejable disponer de dos versiones de materiales distintas: una accesible al público, pensada para su versionado y la recogida de datos de uso (Ferran *et al.*, 2007) y otra reservada a los gestores para su preservación a largo plazo, si bien ambas están enlazadas mediante metadatos. Las versiones para la preservación a largo plazo deberían migrar para que sean accesibles antes de que se conviertan en tecnológicamente obsoletas, tan pronto como se detecte cualquier problema con la versión accesible. Así pues, antes de realizar una ingesta se debe crear una descripción del for-

mato de fichero y almacenarla en los metadatos relativos al formato, intentando separar al máximo las descripciones relativas al contenido de las que hacen referencia a otras propiedades significativas, como por ejemplo la apariencia. Este proceso podría automatizarse en parte, ya que la intervención manual representa un coste elevado y no es sostenible para grandes volúmenes de recursos.

4.2. Evaluación de formatos para la preservación a largo plazo

Finalmente, otro de los aspectos que deben considerarse es la sostenibilidad para los formatos de ficheros. En el caso de la UOC, los materiales se han creado digitalmente pero pensando en una versión en papel. Esta situación muestra que es posible disponer de una gran variedad de formatos y que los materiales más importantes (desde el punto de vista docente) se crean en formato PDF, lo que es aceptable para los estudiantes pero limita su manipulación. No obstante, durante el proceso de ingesta de los objetos de aprendizaje en el repositorio de la UOC, se ha observado que es necesario llevar a cabo algunas migraciones de formato debido a la gran variedad de documentos, para simplificar las necesidades de preservación y también reducir los requerimientos tecnológicos, de acuerdo con los criterios siguientes (no exhaustivos):

Portable Document File: PDF/A, se basa en la PDF Reference Version 1.4. ISO 19005-1:2005, considerada un estándar.

Microsoft Word: transformación a PDF/A u OpenOffice. Hay que tener en cuenta que el fichero puede contener macros y/o formularios.

Microsoft Excel: transformación a XML u OpenOffice.

Vídeo: migración de los ficheros a MPEG2, formato AVI o formato Quick Time.

Documentos LaTeX: conservación del formato original pero a largo plazo DocBook XML con MathML embebido.

Imágenes: conversión de las imágenes a PNG, TIFF o JPEG.

Ficheros PowerPoint: transformación a PDF/A u OpenOffice. Las secuencias interactivas deberían documentarse.

Fichero de audio: transformación a WAV o MP3.

Ficheros Minitab: transformación a ficheros XML.

Ficheros SPSS: formato portable SPSS (*.por) o hacia XML.

Así pues, para poder determinar las prioridades sobre los formatos nos hemos basado en la lista de posibles indicadores de riesgos facilitada por DRAMBORA. Estos indicadores, siendo no excluyentes, son principalmente los siguientes:

- R11 (fallos en la preservación de las características de la información digital)
- R31 (incompatibilidad del software)
- R33 (obsolescencia del software)
- R34 (obsolescencia de los medios de almacenamiento)
- R66 (las estrategias de preservación dan como resultado pérdida de información)

En nuestro caso, estos indicadores permiten establecer las prioridades a la hora de emplear la migración de formatos debido a su obsolescencia tecnológica o su versionado. También nos han permitido determinar qué riesgo es aceptable en la interpretación del riesgo.

Una vez evaluados los posibles riesgos respecto a la obsolescencia tecnológica, ha sido posible establecer una relación entre los procesos de migración y el coste que este supone.

Como hemos indicado previamente, en las operaciones de migración de formatos el coste de transformar un fichero de Microsoft Word a PDF/A difiere significativamente de las operaciones de migración de ficheros de vídeo.

En el proceso de migración de vídeo se pueden producir otros cambios sustanciales en las propiedades significativas, como pérdida de resolución, deterioro del audio o el tiempo que se establece en la operación de migración.

5. Conclusiones

Este artículo ha presentado un análisis de las necesidades de conservación de los objetos de aprendizaje depositados en un repositorio educativo abierto con materiales de estadística creado por la Universitat Oberta de Catalunya. Este repositorio está basado en la plataforma DSpace y promueve la reutilización de materiales educativos así como la preservación a largo plazo de los mismos.

Previamente a las operaciones de preservación digital se han establecido prioridades de operación basadas en la metodología de análisis de riesgos en repositorios como DRAMBORA.

DRAMBORA nos ha permitido establecer prioridades en el plan de preservación teniendo como fin la obsolescencia tecnológica. Como resultado del análisis hemos podido establecer una relación sobre los costes y las operaciones de migración, para poder decidir con que materiales se pueden realizar operaciones de migración que no supongan un cambio importante en sus propiedades significativas.

La amplia gama de formatos usados actualmente para almacenar objetos de aprendizaje implica establecer políticas de preservación a largo plazo para asegurar su recuperación futura, dada la gran variabilidad tecnológica y las necesidades específicas de etiquetado de los contenidos usados en un entorno virtual de aprendizaje, diferentes a los de una biblioteca digital o repositorio institucional. Es necesario, por lo tanto, optar por formatos que aseguren un cierto nivel de continuidad y que faciliten las políticas de preservación, como PDF para documentos y XML para ficheros de datos, aunque siguen existiendo formatos (vídeo, imágenes, etc.) que necesitan ser descritos mediante metadatos adicionales para su posterior recuperación y transformación en caso de que esto sea necesario a causa de su obsolescencia. La utilización de formatos de fichero abiertos basados en herramientas *open source* permite también un cierto nivel de preservación mediante técnicas de emulación, siempre que sea posible compilar y ejecutar el código fuente existente. En este sentido, la existencia de un paquete como OpenOffice ofrece la combinación de ambos elementos, descripción mediante XML y la posibilidad de acceder al código fuente.

En la actualidad, el principal reto pasa por definir mecanismos semiautomáticos que permitan etiquetar y procesar correctamente los miles de objetos de aprendizaje de los que dispone la institu-

ción sin que esto suponga un coste muy elevado imposible de asumir. Igualmente, la introducción de políticas de conservación pasa por retocar aspectos organizativos de la institución que implican a colectivos muy diferentes con objetivos dispares, lo que obligará a establecer flujos de trabajo complejos, teniendo en cuenta la posibilidad de que otros actores adopten el rol de productor de contenidos. Por otra parte, como en cualquier plan de contingencia, es necesario establecer calendarios con simulacros para evaluar periódicamente la obsolescencia de los contenidos almacenados en el repositorio. Finalmente, se espera que en un futuro próximo se haga pública la política de preservación del repositorio institucional.

Agradecimientos

Este trabajo ha sido parcialmente financiado por los proyectos PERSONAL(ONTO), ref. TIN2006-15107-C02, y E-MATH++, ref. EA2008-0151.

Bibliografía

- ASHLEY, Kevin; DAVIS, Richard; PINSENT, Ed (2008). *Significant Properties of E-Learning Objects* [en línea]. [Fecha de consulta: 20 de noviembre de 2010].
<http://www.jisc.ac.uk/media/documents/programmes/preservation/spelos_report.pdf>
- BARKER, Ed [et al.] (2004). *Long-Term Retention and Reuse of E-Learning Objects and Materials* [en línea]. [Fecha de consulta: 20 de noviembre de 2010].
<<http://www.jisc.ac.uk/whatwedo/programmes/preservation/elo.aspx>>
- CCSDS (2002). *Reference Model for an Open Archival Information System (OAIS). Blue Book* [en línea]. [Fecha de consulta: 20 de noviembre de 2010].
<<http://public.ccsds.org/publications/archive/650x0b1.pdf>>
- CHEN, Su-Shing (2004). «Digital Preservation and Workflow Process». En: *Proceedings of the 7th International Conference on Asian Digital Libraries. Digital Libraries: International Collaboration and Cross-fertilization*. Vol. 3.334, págs. 61-72.
- CONWAY, Paul (2008). «Modeling the digital content landscape in universities». *Library Hi Tech*. Vol. 26, n.º 3, págs. 342-354.
- Digital Curation Centre and DigitalPreservationEurope (2007). *DCC and DPE Digital Repository Audit Method Base on Risk Assessment v1.0*. [Fecha de consulta: 20 de noviembre de 2010].
<<http://www.repositoryaudit.eu/download>>
- FERRAN, Núria [et al.] (2009). «User Centered Design of a Learning Object Repository». En: *Proceedings of the 13th International Conference on Human-Computer Interaction. Lecture Notes in Computer Science*. Vol. 5.619, págs. 679-688.
- FERRAN, Núria [et al.] (2007). «Enriching e-learning metadata through digital library usage analysis». *The Electronic Library*. Vol. 25, n.º 2, págs. 148-165.

- Van der HOEVEN, Jeffrey; van DIESEN, Raymond; van der MEER, Kees (2005). «Development of a Universal Virtual Computer for long-term preservation». *Journal of Information Science*. Vol. 31, n.º 3, págs. 196-208.
- LEE, Kyong-Ho [et al.] (2002). «The State of Art and Practice in Digital Preservation». *Journal of Research of the National Institute*. Vol. 107, n.º 1, págs. 93-106.
- MCLEOD, Rory (2008). «Risk Assessment; using a risk based approach to prioritise handheld digital information». En: *Proceedings IPRESS 2008*.
<http://www.bl.uk/ipres2008/presentations_day1/20_McLeod.pdf>
- NEVEN, Filip; DUVAL, Erik (2002). «Reusable learning objects: a survey of LOM-based repositories» [en línea]. En: *Proceedings of the Tenth ACM international conference on Multimedia*. Págs. 291-294. [Fecha de consulta: 20 de noviembre de 2010].
<<http://www.cs.kuleuven.ac.be/cwis/research/hmdb/publications/files/Lorsurvey.pdf>>
- SHREEVES, Sarah; CRAGIN, Melissa (2008). «Introduction: Institutional Repositories: Current State and Future». *Library Trends*. Vol. 57, n.º 2, págs. 89-97.
- SMITH, MacKenzie (2005). «Exploring Variety in Digital Collections and the Implication for Digital Preservation». *Library Trends*. Vol. 54, n.º 1, págs. 6-15.
- WAUGH, Andrew; WILKINSON, Ross; HILLS, Brendan (2000). «Preserving Digital Information Forever». En: *Proceedings of the Fifth ACM Conference on Digital Libraries*. Págs. 175-184.

Sobre los autores

Juanjo Boté

juanjo.botev@ub.edu

Departamento de Biblioteconomía y Documentación de la Universidad de Barcelona

Profesor asociado al Departamento de Biblioteconomía y Documentación de la Universidad de Barcelona desde el año 2009. En el año 2008 participó en el concurso internacional Digital Preservation Challenge donde obtuvo el segundo premio. Ha participado en varias conferencias internacionales vinculadas a la conservación digital.

Departamento de Biblioteconomía y Documentación

Universidad de Barcelona

C/ Melcior de Palau, 140

08014 Barcelona

España

Julià Minguillón

jminguillona@uoc.edu

Universitat Oberta de Catalunya

Julià Minguillón es doctor por la Universidad Autónoma de Barcelona desde septiembre de 2002. En el año 2001 se incorpora como profesor de la Universitat Oberta de Catalunya donde realiza tareas docentes en las áreas de Programación, Lenguajes y Compiladores. Ha participado en la realización de recursos docentes sobre programación orientada al objeto, estructura de la información y tipos abstractos de datos y compiladores. Desde noviembre de 2006 hasta noviembre 2008 ha sido el director adjunto del Internet Interdisciplinary Institute de la UOC. Actualmente es el director académico de la Cátedra UNESCO en e-Learning de la UOC. Sus intereses de investigación incluyen la descripción y estandarización del proceso de aprendizaje mediante ontologías, los objetos de aprendizaje para la personalización mediante itinerarios formativos y el modelado del comportamiento de los usuarios de un entorno de aprendizaje virtual. Ha sido el organizador de la segunda edición del Simposio sobre Objetos de Aprendizaje, SPDECE 2005, y del Seminario Internacional de la Cátedra UNESCO en e-Learning desde su tercera edición. Actualmente ha finalizado la dirección del proyecto PERSONAL financiado por el Ministerio de Educación y Ciencia sobre personalización del proceso de aprendizaje mediante itinerarios adaptativos usando objetos de aprendizaje reutilizables y ontologías, que tiene continuación en el proyecto MAVSEL, sobre minería, análisis y visualización de datos basados en modelos sociales de e-learning. También ha dirigido el proyecto OLCOS financiado por la Unión Europea sobre recursos educativos en abierto y ahora impulsa el proyecto #metaOER, una colección de recursos abiertos sobre recursos educativos en abierto.



Los textos publicados en esta revista están sujetos –si no se indica lo contrario– a una licencia de Reconocimiento 3.0 España de Creative Commons. Puede copiarlos, distribuirlos, comunicarlos públicamente y hacer obras derivadas siempre que reconozca los créditos de las obras (autoría, nombre de la revista, institución editora) de la manera especificada por los autores o por la revista. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by/3.0/es/deed.es>.