

UNIVERSIDAD DE SALAMANCA
FACULTAD DE TRADUCCIÓN Y DOCUMENTACIÓN
Máster en Sistemas de Información Digital
Trabajo de Fin de Máster

Preservación digital en
repositorios institucionales
GREDOS

Alumna: Tránsito Ferreras Fernández

Tutor: José Antonio Merlo Vega

Salamanca, 2010

RESUMEN Y PALABRAS CLAVE

Resumen:

Preservación digital en repositorios institucionales: Gredos es el título del Trabajo Fin de Máster que se enmarca dentro del Máster en Sistemas de Información Digital de la Universidad de Salamanca, curso académico 2009-2010.

En los últimos siete años dos acontecimientos relacionados como son el movimiento Open Access y la creación de los repositorios institucionales han sido detonantes para que el tema de la preservación sea uno de los más estudiados en el ámbito de la documentación digital.

Este estudio trata sobre la preservación digital de los repositorios institucionales en acceso abierto, es decir la preservación digital dirigida a asegurar la accesibilidad y el uso a largo plazo de los objetos digitales depositados en esos repositorios y por lo tanto garantizar la accesibilidad en el futuro a estos últimos.

Además se pretende reflejar el estado de la cuestión actual de la preservación digital y ofrecer una aplicación práctica para *Gredos*, el sistema de Gestión del Repositorio Documental de la Universidad de Salamanca. El objetivo último es, por lo tanto, establecer el *modus operandi* en el desarrollo y puesta en práctica de un plan de preservación digital para *Gredos*.

Palabras clave:

Preservación digital, Políticas, Recursos, Aspectos tecnológicos, Metadatos de preservación, PREMIS, METS, OAIS, Repositorios institucionales, Gredos, Universidad de Salamanca (España).

SUMARIO

Objeto, justificación y metodología	5
Repositorios institucionales	7
Definición	7
Interoperabilidad, recolección, difusión y accesibilidad	8
Beneficios de los RIs	10
Preservación digital	12
Definición y objetivos	12
Amenazas y estrategias	13
Preservación del contenido de los RIs	14
Recursos sobre preservación digital	18
Plan de Preservación digital de GREDOS	30
Gredos	30
Principios de un plan de preservación	32
Factores tecnológicos	35
Creación de recursos digitales en Gredos	35
Sistema y preservación digital con DSpace	36
Metadatos de preservación	38
Premis	38
Mets	42
Uso de Premis con Mets	44
Jhove: extracción automática de metadatos	46
Almacenamiento, soporte y seguridad	47
Formatos	48
Accesibilidad	50
Soluciones frente a las amenazas	51

Conclusiones	54
Bibliografía	56

OBJETO, JUSTIFICACIÓN Y METODOLOGÍA

En los últimos siete años dos acontecimientos relacionados como son el movimiento Open Access y la creación de los repositorios institucionales han sido detonantes para que el tema de la preservación sea uno de los más estudiados en el ámbito de la documentación digital.

Las bibliotecas universitarias españolas no han sido ajenas a este movimiento y prácticamente la totalidad de ellas tienen en marcha o en proyecto su repositorio institucional. Los datos sobre el número de repositorios institucionales en España ascienden a 51, según el directorio *DOAR* <http://www.opendoar.org/>. De estos 51 repositorios institucionales, 43 pertenecen a universidades y centros de investigación, según los datos de *RIs Busca Repositorios* <http://www.accesoabierto.net/repositorios/>.

Así la mayoría de las bibliotecas universitarias españolas están embarcadas en proyectos de digitalización con fines de difusión, depósito y preservación digital. En todas las universidades, además de recoger, almacenar y difundir información digital, se pretende que esa información perdure, que esté disponible y que sea recuperable a lo largo del tiempo sin que se vea afectada por los inevitables cambios tecnológicos producidos a velocidad vertiginosa. Por ello, es importante que cada institución comprometida en proyectos de digitalización tenga en cuenta el aspecto de la preservación a largo plazo que permita el acceso, la recuperación y la consulta de la información digital y de sus objetos digitales.

Cuando se habla de preservación digital hay que referirse a la conservación para asegurar la accesibilidad y la recuperación de materiales digitales. También puede entenderse como preservación digital, la técnica de digitalización de documentos en soporte papel para asegurar su conservación en el tiempo y evitar el deterioro del original. En este estudio, se tratará la preservación digital de los repositorios institucionales en acceso abierto, es decir la preservación digital dirigida a asegurar la accesibilidad y el uso a largo plazo de los objetos digitales depositados en esos repositorios y por lo tanto garantizar la accesibilidad en el futuro a estos últimos.

En este trabajo se pretende reflejar el estado de la cuestión actual de la preservación digital que nos permita llegar a ofrecer una aplicación práctica para Gredos, el sistema de Gestión del Repositorio Documental de la Universidad de Salamanca. El objetivo último es, por lo tanto, establecer el *modus operandi* en el desarrollo y puesta en práctica de un plan de preservación digital para Gredos, el repositorio institucional de la Universidad de Salamanca.

El trabajo se presenta dividido en tres capítulos. En el primero de ellos se hace una breve introducción sobre lo que son los repositorios institucionales, la razón de ser de éstos, y las funciones y funcionalidades de la interoperabilidad, recolección, difusión y accesibilidad de la información digital. Los beneficios que los Repositorios Institucionales (RIs) aportan al mundo de la información científica es motivo suficiente para que las instituciones afronten como una prioridad la preservación digital de sus repositorios.

El segundo capítulo además de definir y centrar el tema del trabajo, la preservación digital, trata de ofrecer un estado del panorama mundial sobre preservación a través de los recursos existentes y las estrategias aplicadas en este campo. Se realiza un resumen recopilatorio de recursos existentes en este campo que pueden servir de guía a la hora de estudiar la actualidad sobre la preservación digital.

El último capítulo trata sobre Gredos en general, y sobre su preservación digital en particular, sobre los principios básicos a tener en cuenta a la hora de plantearse una política de preservación digital en un repositorio institucional como Gredos, y sobre la aplicación práctica de un plan de preservación a partir de las experiencias de otras instituciones en este campo. Se analizan los factores técnicos imprescindibles de un plan de preservación a largo plazo, exponiendo a la vez qué elementos técnicos se han implementado ya en este repositorio y cuáles son sus carencias, haciendo hincapié en los aspectos que aún no están recogidos y que se deben desarrollar para la puesta en marcha de un plan de preservación a largo plazo. En relación con Gredos se estudia la aplicación de la tecnología que ofrece DSpace, software de gestión del repositorio. Por último, se plantean las posibles amenazas a las que puede verse sometido el repositorio y las posibles soluciones que hay que prever para no perder la información ni la accesibilidad a esta.

REPOSITARIOS INSTITUCIONALES

DEFINICIÓN

Un repositorio institucional (RI) es un conjunto de servicios para almacenar y hacer accesibles materiales de investigación en formato digital creados por una institución y su comunidad, una colección digital del producto de la investigación llevada a cabo por esa comunidad. Los repositorios institucionales se están convirtiendo en herramientas esenciales para la comunicación académica en la era digital. Ellos pueden formar parte de un sistema mayor, nacional, regional y global de repositorios, indizados de una manera estándar y recuperable, utilizando una interfaz de acceso. También pueden proveer las bases para nuevos modelos de edición académica y pueden vincularse a otros servicios relacionados, como enseñanza a distancia electrónica, publicación electrónica de revistas de investigación o publicación por demanda. (Lynch, 2003)

Las definiciones que se han dado de repositorio institucional han sido numerosas. Sánchez y Melero (Sánchez and Melero, 2006) han hecho una recopilación de las mismas. Entre ellas, las siguientes:

- SPARC (Crow, 2002) lo define de forma muy escueta como una colección digital que agrupa y preserva la producción intelectual de una o varias universidades. Establece un objeto de aplicación muy amplio, el conjunto de la producción intelectual de la institución, en el que cabría más allá de los resultados de investigación y objetos de aprendizaje cualquier otro producto de carácter cultural producido por la institución.
- Clifford Lynch (Lynch, 2003) define un repositorio institucional como un conjunto de servicios que una universidad ofrece a los miembros de su comunidad para la gestión y diseminación de los materiales digitales creados por la institución y sus miembros. Destacable la consideración de servicio que se da al repositorio, como algo dinámico al servicio de la comunidad, más allá de un depósito pasivo de documentos digitales.
- Cat S. McDowell (McDowell, 2007) considera que un repositorio debe cumplir al menos las siguientes características:
 - Es un servicio institucional abierto a toda la comunidad universitaria y a todo tipo de temáticas.
 - Su objetivo debe ser reunir, preservar y dar acceso a, entre otras cosas, la producción de los investigadores y docentes en múltiple formatos. Se excluiría cualquier repositorio que ponga límites a esta producción, por ejemplo, sólo objetos de aprendizaje o sólo de tesis, etc.
 - Debe recibir contenidos de forma activa bien a través de un formulario web o simplemente a través de correo electrónico.

De estas definiciones extraemos en conclusión que un repositorio institucional es un conjunto de servicios prestados por las universidades, al conjunto de la comunidad, para recopilar, administrar, difundir y preservar la producción documental digital generada en la institución, cualquiera que sea su tipología, a

través de la creación de una colección digital organizada, abierta e interoperable a través del protocolo OAI-PMH (protocolo para la recolección de metadatos) para garantizar un aumento de la visibilidad e impacto de la misma.

INTEROPERABILIDAD, RECOLECCIÓN, DIFUSIÓN Y ACCESIBILIDAD

Entendemos por interoperabilidad la capacidad de los sistemas, servicios y organizaciones para trabajar juntos de un modo transparente hacia objetivos comunes o diferentes. En el ámbito técnico se apoya, entre otros, en los estándares abiertos para la comunicación entre sistemas y para la descripción de recursos y colecciones. La interoperabilidad tiene que ver principalmente con el contexto de la búsqueda y acceso a los recursos. La Iniciativa de Archivos Abiertos (OAI) pretende desarrollar y promocionar normas de interoperabilidad que tienen como objetivo el facilitar la difusión eficaz de los contenidos. El objetivo esencial de los archivos abiertos es permitir el acceso a los materiales en la web por medio de repositorios que, interoperando unos con otros, permitan el intercambio de metadatos, su almacenamiento y publicación. La necesidad de la comunidad de e-print (ediciones preliminares electrónicas) de una solución de interoperabilidad de bajo nivel para el acceso entre repositorios heterogéneos dio lugar a la creación de la Open Archives Initiative (OAI). En la misión de la OAI se dice: "La Open Archives Initiative desarrolla y promueve normas de interoperabilidad que tienen por objeto facilitar la difusión eficaz de los contenidos" (<http://www.openarchives.org/documents/FAQ.html#What%20is%20the%20mission%20of%20the%20Open%20Archives%20Initiative>).

Muchas comunidades se podrán beneficiar de esta estrategia de archivos abiertos. Internet y la gran cantidad de documentos en formato digital han aumentado el número de usuarios potenciales de los repositorios. Los documentos pueden ser más accesibles y además utilizados con distintos objetivos de los que motivaron inicialmente la creación de los repositorios. Por otra parte, la posibilidad de acceder a múltiples repositorios permite la construcción de nuevos tipos de servicios que puedan atender mejor las necesidades de los usuarios.

En relación con la OAI, la recolección se refiere específicamente a la recogida de los metadatos de una serie de repositorios distribuidos en un almacén de datos combinados. Y la aplicación cliente que suministra las peticiones OAI-PMH sería un *recolector* o *harvester*.

El Protocolo OAI para la Recolección de Metadatos (OAI-PMH), define un mecanismo para la recolección de registros que contienen los metadatos de los repositorios. Ofrece a los *proveedores de datos* una opción técnica sencilla para poner sus metadatos a disposición de servicios basados en los estándares abiertos HTTP (Hypertext Transport Protocol) y XML (Extensible Markup Language). Los metadatos que son recolectados pueden estar en cualquier formato establecido por una comunidad (o por cualquier conjunto específico de *proveedores de datos* y *proveedores de servicios*), con independencia de que hayan establecido el Dublin Core no cualificado para proporcionar un nivel básico de interoperabilidad. Así pues, los metadatos de

distintas fuentes pueden ser reunidos en una base de datos, y se pueden ofrecer servicios sobre la base de esta recolección centralizada, o "agregación" de datos. El vínculo entre estos metadatos y el contenido no está definido por el protocolo OAI. El protocolo OAI-PMH no proporciona una búsqueda a través de estos datos, simplemente permite reunir los datos en un sitio. Para ofrecer un servicio, el método de la recolección debe combinarse con otros mecanismos.

Es conveniente aclarar los conceptos *proveedores de datos* y *proveedores de servicios*. Los *proveedores de datos* manejan el depósito y la publicación de los recursos en un repositorio y exponen los metadatos de los recursos del repositorio para que puedan ser recolectados. Ellos son los creadores y conservadores de los metadatos y de los repositorios de recursos. Los *proveedores de servicios* recolectan los metadatos de los *proveedores de datos* y, emplean los metadatos recolectados con el fin de proporcionar servicios. El tipo de servicios que se pueden ofrecer son: una interfaz de búsqueda, un sistema de evaluación por pares, etc. Hay que tener en cuenta que una organización proveedora puede desempeñar ambas funciones, tanto la de proporcionar datos para su recolección como la de proporcionar servicios para el usuario final.

El uso del protocolo es muy prometedor dentro de la estrategia de los archivos abiertos. Podríamos citar como beneficios más conocidos el servir de sustento de un nuevo modelo de comunicación académica, y el hacer aflorar a la superficie los recursos ocultos y la interoperabilidad a bajo coste.

Podemos afirmar que OAI-PMH es uno de los mayores aciertos en el acceso a bases de datos bibliográficas (y otros datos) en los últimos diez años. Gracias a su sencillez y facilidad de aplicación, a lo que contribuye el hecho de que el esquema básico para los metadatos sea Dublin Core (ISO 15836:200945) se ha extendido enormemente, empujando, y recibiendo a su vez el empuje, de movimientos como Open Access (OA) y dando lugar a "repositorios abiertos", en el sentido de los principios de difusión de contenidos que animan este movimiento, pero también en el sentido de ser susceptibles de ser recolectados, como se ha venido explicando anteriormente.

Los repositorios institucionales han sido los más beneficiados de la rápida expansión del protocolo y de las aplicaciones basadas en él. El retraso en la aplicación de la normativa OAI por parte de las empresas de software para bibliotecas ha provocado que los repositorios institucionales se hayan constituido en muchos casos en una biblioteca digital paralela al sistema de gestión bibliotecario utilizado para los materiales bibliográficos.

El gran éxito de los sistemas OAI-PMH puede verse en las cifras que ofrecen algunas aplicaciones. Según *Wayback Machine*, en junio de 2003 *OAIster*, creado en 2002 como un proyecto de la *University of Michigan Digital Library Production Services* contaba con 1.089.937 registros de 142 instituciones. En 2009 OCLC anunció que pasaba a alojar y administrar *OAIster* e incluyó los registros recolectados a través de OAI en *WorldCom*. Las cifras a principios de 2010 son de más de 23 millones de registros procedentes de 1.100 repositorios.

El hecho de que los repositorios dispongan de protocolo OAI hace que sean visibles a través del mayor catálogo del mundo o de ser buscables a través de los grandes buscadores como *Google* o *Yahoo* que indizan directamente *OAIster*. Para ello es preciso registrar el repositorio en diferentes registros, *OAI Registered Data Providers* o en los proveedores de servicios como *OAIster*, *Registry of Open Acces Repositories (ROAR)*, *Recolecta: Recolector de Ciencia Abierta o Hispana*, *Directorio y Recolector de recursos digitales*.

Además, *Hispana* es el agregador español de *Europeana*, la biblioteca digital europea, y *Recolecta* es el agregador de *DRIVER*, iniciativa europea (*Digital Repository Infrastructure Vision for European Research*) que tiene como objetivo establecer una infraestructura cohesionada y paneuropea de repositorios digitales y ofrecer servicios y funcionalidades para investigadores y para el público en general. A través de *DRIVER* se recolectan trabajos científicos, informes científicos y técnicos, informes de trabajo, pre-prints, artículos y datos, y es a su vez uno de los agregadores especializados de *Europeana* a la que aporta las publicaciones científicas.

Tanto para participar en *Europeana* como para participar en *DRIVER* los repositorios deben cumplir unas directrices de aplicación, además del protocolo OAI-PMH, para la adecuación de éstos a la metodología y normativa específica de estos dos recolectores. En *Europeana* se trata de los elementos ESE, *Europeana Semantic Elements* (www.europeana.eu/schemas/ese/ESE-V3.2.xsd), y en *DRIVER* de las *Directrices para proveedores de contenido* (www.driver-support.eu/.../DRIVER_2_0_Guidelines_Spanish.pdf).

BENEFICIOS DE LOS RIs

Implementar un repositorio institucional requiere un considerable esfuerzo de planificación y de compromiso. Entonces, ¿por qué tantas Universidades están creándolos?

Una de las principales motivaciones es la de permitir el acceso abierto a los resultados de su actividad científica y académica. Pero quizás esta razón solamente sería insuficiente para justificar el esfuerzo considerable que supone la creación de repositorios.

En un primer nivel, un repositorio institucional es el reconocimiento de que la actividad intelectual de nuestra universidad estará representada cada vez más en soporte digital y que la principal responsabilidad de esta es ejercer el control sobre su producción intelectual, haciéndola accesible, fácilmente recuperable y asegurando su permanencia en el tiempo.

Las Instituciones de Educación Superior tienen que gestionar sus contenidos digitales de forma más efectiva y transparente que antes. En Europa se están desarrollando actividades (EEES) que ofrecerán beneficios tangibles a aquellas instituciones que puedan demostrar y explotar sistemas y estrategias de información efectivas.

La gran ventaja que proporcionan estos sistemas es que ayudan a la institución a desarrollar estrategias coherentes y coordinadas para la captura, identificación, almacenamiento, conservación y recuperación de sus contenidos

digitales. El tratamiento gestionado de estos contenidos aumenta las oportunidades para un uso más efectivo de los resultados de la actividad de la institución y estimula la colaboración entre las diferentes disciplinas y comunidades. La posibilidad que ofrecen estos repositorios de reutilizar los contenidos abre un amplio abanico de aplicaciones, por ejemplo en el ámbito de los objetos de aprendizaje. Por otra parte, suponen un medio de romper el ciclo de depósitos o almacenes individuales de conocimiento dentro de la institución, ofreciendo un espacio común de almacenamiento con acceso para todos.

Un repositorio institucional ofrece la difusión más amplia posible de toda la oferta de la producción intelectual digital generada en una institución, incrementando la visibilidad y el prestigio de la misma y demostrando su valor para las fuentes de financiación y subvenciones. Además puede aumentar la visibilidad de sus investigadores, ampliando la difusión y el uso de sus trabajos; puede estimular la innovación, facilitar un análisis cualitativo del trabajo de sus miembros, apoyar las tareas de enseñanza y aprendizaje, servir de sistema de registro de ideas y ofrecer un catálogo del capital intelectual de la institución. Los repositorios institucionales reconocen y dan cabida a los objetos digitales científicos no incluidos en los canales tradicionales de publicación. Finalmente, los repositorios institucionales pueden mejorar la comunicación científica y hacer avanzar la investigación permitiendo a los usuarios localizar y recuperar información relevante más rápida y fácilmente.

Ante esta realidad es una prioridad de las instituciones, en este caso de la Universidad, tomar conciencia de la necesidad de conservar y hacer accesible a largo plazo el contenido digital de los repositorios institucionales, como reflejo que son de la actividad investigadora y docente de la Universidad, a la vez que depositarios del fondo patrimonial de la misma.

PRESERVACIÓN DIGITAL

DEFINICIÓN Y OBJETIVOS

La preocupación por salvaguardar la información digital y por el acceso permanente a ella a largo plazo es lo que hace que se haya tomado conciencia de un problema al que hay que encontrarle una solución que nos asegure la permanencia y la accesibilidad a la avalancha de información digital que se está produciendo actualmente.

Ninguna generación había producido antes tantos documentos como la nuestra. La mayoría de los testimonios que dejamos son digitales. Nuestra preocupación es si esta documentación llegará a las generaciones futuras. Si además unimos a esto el hecho de que los soportes son frágiles y que los cambios en las tecnologías son frecuentes, el panorama no es muy alentador. Neil Beagrie (2004) afirma que la información digital nunca sobrevivirá accidentalmente.

En los últimos siete años dos acontecimientos relacionados como son el movimiento Open Access y la creación de los repositorios institucionales han sido detonantes para que el tema de la preservación sea uno de los más estudiados en el ámbito de la documentación digital.

Pero cabe preguntarnos qué es la preservación digital y en qué fase nos encontramos en este terreno. ¿Podemos asegurar que dentro de treinta años seguiremos teniendo acceso a la documentación digital si aplicamos las técnicas vigentes hoy en esta materia?

En las *Directrices para la preservación digital* (UNESCO, 2003) se define la preservación como las acciones destinadas a mantener la accesibilidad de los objetos digitales a largo plazo.

En *DPC Handbook* (2008) se define como las actividades necesarias para asegurar el acceso continuado a materiales digitales hasta cuando sea necesario, a pesar de los obstáculos que representan los fallos en los soportes o los cambios tecnológicos.

En *Trusted digital repositories* (2002) son las actividades gestionadas necesarias para asegurar el mantenimiento a largo plazo de la cadena de bits y la accesibilidad continuada del contenido.

Waller y Sharpe (2006) dicen que son las acciones necesarias para asegurar un acceso continuado y pertinente a la información digital durante el tiempo que sea requerido y para cualquier finalidad legítima.

La preservación digital se define como un conjunto de procesos dirigidos a conservar la información en formato digital. No existe preservación digital si no se mantiene la posibilidad de acceder a los recursos digitales. El objetivo de la preservación digital es permitir a los futuros usuarios recuperar, acceder, descifrar, ver, interpretar, entender y experimentar documentos y datos de forma significativa y válida (J. Rothenberg, 1995).

Como vemos a través de las múltiples definiciones de los autores, la preservación digital siempre va encaminada a asegurar el acceso continuado a la información digital a largo plazo.

En la *Carta para la preservación del patrimonio digital (UNESCO, 2003)* se expone la necesidad de salvaguardar el patrimonio digital, de tomar medidas para su preservación y de asegurar el acceso a él. El objetivo de la conservación del patrimonio digital es que este sea accesible para el público.

En ese mismo documento se contemplan los riesgos de la documentación digital y se afirma que el patrimonio digital del mundo corre el peligro de perderse para la posteridad. Entre otros factores que contribuyen a ello, se encuentran la rápida obsolescencia de los equipos y programas informáticos, las incertidumbres existentes en torno a los recursos, la responsabilidad y los métodos para su mantenimiento y conservación, y la falta de legislación que proteja estos procesos.

Hoy en día, los repositorios institucionales forman parte de ese patrimonio digital al que hace referencia la Carta de la Unesco en 2003. Se hace necesario, llegado este punto, poner en práctica estrategias de preservación digital. La preservación digital representa *“un problema vital para una sociedad de la información en la que la oferta de información crece exponencialmente y aumenta sin cesar el dinamismo de los contenidos. [...] Es un problema que tanto los políticos como las instituciones más afectadas deben acometer con urgencia”* (Comisión Europea, 2005).

AMENAZAS Y ESTRATEGIAS

Los repositorios institucionales en tanto que elementos de la sociedad de la información ya están formados y funcionando desde hace unos años, los esfuerzos deben dirigirse ahora a afianzar y a asegurar el acceso a la información contenida en ellos en un futuro cercano y a largo plazo. Nos enfrentamos a amenazas que hay que tener en cuenta a la hora de pensar en la preservación digital:

- los objetos digitales son inestables,
- los equipos y los programas tienen una vida media de unos pocos años,
- pérdidas por desastres naturales, accidentes, averías, virus...,
- barreras al acceso: claves, cifrado, acceso restringido,
- responsabilidades poco claras,
- falta de conocimientos o de recursos,
- problemas legales: copyright,
- descripción inadecuada: imposible recuperación,
- pérdida de información sobre el contexto.

Las principales estrategias aplicadas actualmente para frenar la pérdida de información digital, debida a las causas que se han expresado anteriormente, son principalmente:

- preservación de la tecnología
- migración de los datos
- emulación de las aplicaciones informáticas originales

Pero, según el informe *Digicult report: technological landscapes for tomorrow's cultural economy; unlocking the value of cultural heritage* (Office for Official Publications of the European Communities, 2002) estas estrategias son soluciones inmediatas para un problema a largo plazo. Es decir, nuestros conocimientos actuales más punteros no nos garantizan la capacidad de preservar lo que estamos creando para un futuro medio y lejano. <http://digicult.salzburgresearch.at/>.

No se trata solo de problemas técnicos, la razón por la que nuestros esfuerzos para preservar el objeto digital no son suficientes es que existen retos más allá de los tecnológicos que representan obstáculos más complejos que hay que superar. Son problemas de distinta índole:

- Legales: si no tenemos el permiso del titular no podremos preservar un recurso reproduciéndolo o reformateándolo.
- Económicos: si no contamos con los medios necesarios no podremos garantizar la perdurabilidad de los documentos a lo largo de los años.
- Institucionales: si no se asegura el compromiso institucional permanente, si no están convencidos todos los implicados de la necesidad de colaborar, si no podemos verificar si se cumple todo lo que se promete, no podremos preservar a largo plazo.

La comunicación de la Comunidad Europea sobre la iniciativa *i2010: bibliotecas digitales* (2005) expone una situación que seguía tan precaria como la observada en el *Digicult report*, tres años antes: "Actualmente la experiencia con que se cuenta [...] es escasa, el marco jurídico inestable, los recursos escasean y el resultado de los esfuerzos de preservación es incierto".

De todas formas, ha habido claros progresos en algunas áreas. En los últimos años muchas instituciones internacionales y también en España se han comprometido a crear repositorios digitales, sobre todo en el entorno universitario donde se está implantando software especializado a este fin, como por ejemplo *Dspace*, desarrollado por el Massachusetts Institute of Technology (MIT) o *Eprints*, producto de la universidad británica de Southampton.

El principal objetivo de estos repositorios ha sido el asegurar la captación y la descripción de los recursos digitales de su institución para su recuperación posterior, de acuerdo a la filosofía del acceso abierto. Pero son pocas las instituciones que hasta ahora han afrontado plenamente el problema de la preservación a largo plazo.

LA PRESERVACIÓN DEL CONTENIDO DE LOS RIs

El material contenido en los repositorios institucionales es material objeto de la preservación. La clase de documentos allí depositados puede ser muy variada:

- resultados de la investigación (tesis, e-prints);
- objetos de docencia y aprendizaje;
- datos no elaborados;
- fondos digitalizados;
- material administrativo;
- fondo patrimonial.

De este material cabe preguntarse cuál es el que se debe conservar a largo plazo. La teoría dice sobre esta relación preservación / repositorio institucional que: *Un compromiso con la administración para la organización de estos materiales digitales, incluye la preservación a largo plazo en su caso, así como la organización y el acceso o distribución* (Lynch, 2003)... *La mayoría incluiría la preservación a largo plazo de materiales digitales como una función clave de los Repositorios Institucionales* (Pathfinder report, 2004)... *Los RIs, en nuestra opinión, serán un vehículo importante para hacer frente a estas obligaciones preservación de datos* (Lynch/Lippincott, 2005)... *Conservación y fiabilidad son sin duda el fundamento de cualquier repositorio* (Chavez et al., 2007)...

En cambio la realidad nos muestra que, por ejemplo, “*el Repositorio Institucional de Cranfield University está más preocupado por el acceso que por la preservación*” (Bevan, 2007) y que según un estudio realizado por Hitchcock “*ninguno de los repositorios estudiados tiene una política formal de preservación para guiar las decisiones sobre cuestiones planteadas en la encuesta*” (Hitchcock et al., 2007).

Los criterios tradicionales para poner en práctica la preservación a largo plazo tienen que ver con los siguientes factores:

- valor
- pertinencia
- uso
- otros condicionantes a tener en cuenta son: misión, disponibilidad de recursos humanos, económicos, materiales, obligaciones legales o contractuales.

Los nuevos criterios sobre los recursos digitales según el *DPC Handbook* son:

- que la institución tenga pleno derecho a manipular los datos para asegurar su acceso en entornos informáticos del futuro;
- que el recurso sea de un formato legible actualmente y previsiblemente en el futuro;
- que el recurso esté en un soporte gestionable para su transferencia y/o almacenamiento;
- que el recurso disponga de documentación, incluyendo los metadatos.

La selección de recursos para su preservación incluye la decisión sobre qué formatos, qué versiones, qué tipo de material adicional incluir y qué atributos se quieren preservar (datos y funcionalidad, apariencia y esencia). La decisión dependerá de la misión institucional, las necesidades de la comunidad de usuarios, la capacidad técnica/ tecnológica institucional y los recursos disponibles. Los gestores de los repositorios tendrán que decidir sobre qué tipos de archivos aceptarán sus repositorios. Un ejemplo de selección de tipo de formatos es la Universidad de Loughborough (Bates, M. et al., 2006) que usaba los siguientes formatos de archivos:

Texto
<ul style="list-style-type: none"> • Microsoft Word • PDF
Otros de Microsoft Office
<ul style="list-style-type: none"> • Microsoft PowerPoint

• Microsoft Excel
• Microsoft Access
Video/Animation files
• Flash
• QuickTime
Ficheros basado en web
• HTML
Imágenes
• GIF
• JPEG
• BMP
Otros
• CAD
• FileMaker Database
• Hot Potatoes
• LaTeX

Por otra parte, un tema fundamental es el de los derechos de autor. El Repositorio Institucional necesita obtener permiso de los titulares para realizar acciones de difusión, reproducción y modificación / transformación. Además este permiso posiblemente tendrá que extenderse a terceros, por ejemplo proveedores de servicios de preservación.

En cuanto a la utilización de metadatos en repositorios institucionales, una de las primeras medidas de un sistema de preservación digital es la asignación de metadatos a los objetos digitales. Los metadatos son necesarios para permitir:

- el acceso a los documentos depositados
- su gestión y preservación a largo plazo

La clasificación más comúnmente aceptada de metadatos es la siguiente:

1. Los metadatos descriptivos: representan los datos sobre el contenido intelectual, que ayudan a identificar y localizar un recurso. El estándar más utilizado es Dublin Core. Desde sus orígenes en 1995, *Dublin Core Metadata Element Set (DCMES)* ha tenido una gran éxito de implantación. Estas especificaciones técnicas para la descripción de todo tipo de recursos han sido revisadas en diferentes ocasiones y constituyen la norma ISO 15836:2009 (y ANSI Z39.85-2007). Aunque originada en un entorno bibliotecario, con la participación de destacados miembros de la OCLC, estaba pensada para su aplicación, fundamentalmente, fuera del entorno bibliotecario, siendo muy sencilla (15 elementos) y completa para garantizar su aplicación. La versión inicial, 1.0 de 1998 ha tenido muchas modificaciones, hasta llegar a la actual versión de 2010. La norma es muy básica y no incluye apenas reglas de codificación para los valores de determinados elementos, ni guía, salvo un breve comentario, para la redacción de los elementos. A los 15 elementos básicos de DCMES, conocido también como *Dublin Core básico o sin cualificar*, se le añadieron en el año 2000 una serie de cualificadores y de refinamientos de codificación que ampliaban las posibilidades de descripción de los recursos y hacían la norma mucho más completa. En resumen, la especificación de todos los términos que

mantiene *Dublin Core Metadata Initiative* conforman *DCMI Metadata Terms* del que DCMES (o Dublin Core sin cualificar) es un subconjunto. Dublin Core ofrece la ventaja de que al no estar orientado a un dominio específico puede ser la estructura de datos común a la que poder convertir otras estructuras de datos procedentes de repositorios, lo que lo hace especialmente interesante para sistemas de información creados a partir de la integración de metadatos recolectados de fuentes heterogéneas. El sitio oficial de DCMI ofrece los esquemas XML recomendados para la codificación tanto de DCMIES (DC sin cualificar) como de *DCMI Metadata Terms* (DC cualificado), así como el *DCMI RDF schema*, y en HTML/XHTML. Además, DC puede expresarse en el *XML Schema for Dublin Core without qualification*, siguiendo la versión 2.0 de *The Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) y en *Dublin Core Extended*, siguiendo la norma SRU (*Search/Retrieval via URL*). Agenjo y Hernández (2010) remarcan que la prueba de la expansión que ha adquirido DC está en ser la estructura de datos descriptivos utilizada por la mayoría del software para repositorios digitales y en ser el núcleo central de *Open Archives Initiative Protocol for Metadata Harvesting*. Una idea de las cifras que arrojan la combinación DC + OAI está en los más de 23 millones de objetos digitales procedentes de 1.139 repositorios que ofrece OAIster a 1 de marzo de 2010.

2. Metadatos administrativos. Datos necesarios para la gestión de un recurso: mantenimiento, almacenamiento, entrega; incluyen los aspectos técnicos y los relacionados con los derechos. Se trata pues de todos aquellos esquemas que describen la procedencia de un objeto digital, los procesos realizados para su creación o generación, sus características técnicas, sus condiciones de acceso y derechos de propiedad intelectual, y las acciones ya realizadas o previstas relacionadas la preservación del objeto mismo. La codificación de este tipo de datos va encaminada a mantener los objetos digitales utilizables para todos los procesos y funcionalidades propios de una biblioteca virtual, para los que el acceso es el fundamental, pero que afectan también a su gestión y administración.
3. Metadatos estructurales. Se trata de los datos que se refieren a la estructura del recurso y a los elementos que lo integran.

La creación de los metadatos es un trabajo profesional laborioso similar a la catalogación o indización. Pero a diferencia de esas operaciones, los elementos son difíciles de conocer, el volumen de documentos digitales generados es muchísimo mayor y hay un alto porcentaje de documentos únicos, además de existir la necesidad de actualizar los metadatos a lo largo de la vida del documento (Gallart y Keefer, 2008).

La creación de los metadatos puede correr a cargo del personal del repositorio, de los creadores o depositantes, o pueden ser de generación automática a partir del texto o de otras fuentes externas como las bases de datos. La Universidad de Hull a través del proyecto *RepoMMan* recomienda la automatización a la hora de crear los metadatos ya que no se considera viable que los metadatos estén generados enteramente por personas (human-

generated). Se recomienda crear los metadatos según su clase a partir de algunas herramientas (Green et al, 2007):

- Metadatos “contextuales”, extraerlos de los sistemas administrativos institucionales.
- Metadatos descriptivos, extraerlos del documento mediante el programa “iVia” (Univ.California - Riverside).
- Metadatos técnicos, generados por JHOVE.

JISC ha desarrollado algunos proyectos sobre metadatos:

- AIR (Automated Archiving for an Institutional Repository): Universidad de Wolverhampton, septiembre de 2007 – marzo de 2009.
- Digital Plait: Universidad de Aberystwyth, mayo – diciembre de 2008.
- EIDeR: Kings College London, abril de 2008 – marzo de 2009.

El último, el proyecto EIDeR, Enhanced ingest to digital e-research repositories, desarrollado en el Kings College London www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/interoperabilityde/mos/eider.aspx, se trata de desarrollar interfaces “máquina-a-máquina” para la extracción automática de metadatos, la preservación, los derechos de acceso, depósito y captura.

RECURSOS SOBRE PRESERVACIÓN DIGITAL

Los recursos más representativos estudiados se clasifican en tres apartados que responden a la tipología de los recursos:

- Informes.
- Estándares y directrices.
- Proyectos.

INFORMES

Se trata de informes representativos, de ámbito nacional e internacional, redactados a petición de algunas entidades u organismos para la puesta en marcha de planes de preservación digital.

Biblioteca Nacional de Australia (2003). Directrices para la preservación del patrimonio digital.

<http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>.

La información de origen digital (la más democrática de la historia) que se genera en todo el mundo, debe ser preservada de forma que pueda ser utilizada por todos en el presente y en el futuro, ya que es la memoria de la humanidad. Este documento, aunque toma como base el de UNESCO (lo lleva incluso reproducido en sus páginas), es mucho más profundo, ya que trabajan conjuntamente al respecto. Insiste en la necesidad de adoptar unas características técnicas de preservación que perduren en espacio y tiempo, compatibles con toda la tecnología actual y la que ha de venir en un futuro, así como en la claridad de ideas a la hora de decidir qué es lo que debe preservarse. Contiene bibliografía y glosario terminológico muy aclaratorio. Los objetivos del presente documento son:

- Salvar la complejidad electrónica existente en la actualidad y fijar estándares y directrices, formatos, códigos, etc. ya que se ignora cómo va a evolucionar la tecnología (inestabilidad de Internet).
- Hacer que el trabajo que se desarrolle hoy día no vaya a resultar inútil y vayamos a perder patrimonio digital, por ello es necesaria la cooperación (incluso la del trabajo en equipo) y el establecimiento de alianzas.

IFLA (Apéndices A, B y C de la traducción española) (2008).

<http://archive.ifla.org/VII/s19/pubs/digit-guide-es.pdf>

http://travesia.mcu.es/documentos/pautas_digitalizacion.pdf

El apéndice A se trata del Borrador de prescripciones técnicas para digitalizar colecciones de patrimonio bibliográfico y el apéndice B es el Borrador de prescripciones técnicas para digitalizar prensa histórica.

Metadata for Long Term Preservation / by Catherine Lupovici and Julien Masanès, Bibliothèque Nationale de France. - Den Haag : Koninklijke bibliotheek, 2000. - (NEDLIB Report series ; 2).- With summary ISBN 90-62-59-1469

<http://nedlib.kb.nl/results/NEDLIBmetadata.pdf>

Este informe es parte del proyecto europeo sobre bibliotecas nacionales NEDLIB (Networked European Deposit Library) que finalizó en 2001 y tenía como propósito construir una infraestructura básica sobre la que se pudiera llevar a cabo una biblioteca europea de depósito en red. El objetivo de este informe es definir un conjunto mínimo de metadatos que sean obligatorios en cuestiones de gestión de la preservación para procesar grandes cantidades de datos en un entorno tecnológicamente cambiante como es el actual. En él no se abordan otros tipos de metadatos como pueden ser los descriptivos, administrativos o legales. Tampoco trata de cubrir en detalle todos los tipos de documentos digitales con sus particularidades, sino que se limita a la información más general. Propone definir 8 elementos de metadatos y 38 subelementos siguiendo la taxonomía de objetos de información definida en OAIS. Se propone un análisis de la información del documento digital en niveles con objeto de recopilar todos los datos necesarios para el procesamiento de la cadena de bits que lo representa. Los metadatos están pensados para que puedan ser creados, en el mayor número posible, por procedimientos automatizados. Los metadatos propuestos deberían ser utilizados dentro del Sistema de Depósito para ubicaciones Electrónicas de NEDLIB basado en el modelo OAIS. OAIS, como modelo de referencia, cubre todas las funciones implicadas en el conjunto de la cadena de procesos para recopilar los documentos, para proceder a su almacenamiento, para su preservación a largo plazo y para el acceso por aquella comunidad de usuarios que se defina.

Preservació de dipòsits digitals (2007) / Nuria Gallart y Alice Keefer

<http://www.recercat.net/bitstream/2072/4078/1/0703Preservacio.pdf>

El CBUC mantiene actualmente depósitos digitales como TDX, RACO, RECERCAT y Memoria Digital de Cataluña, pero para garantizar una mejora actual y continuar la preservación en el tiempo, va a encargarse de informes que le orienten como proceder (institucional y técnicamente), mediante el estudio de proyectos existentes. Actualmente es necesario el estudio de los recursos

digitales y humanos mediante revisión bibliográfica de todo lo novedoso que actualmente hay publicado al respecto, así como estudios evaluadores.

UNESCO. Carta para la preservación del patrimonio digital (2003)
http://portal.unesco.org/ci/en/files/13367/10676067825Charter_es.pdf/Charter_es.pdf

La preservación a largo plazo del patrimonio digital mundial, para que este no desaparezca, requiere de sistemas y procedimientos fiables que generen objetos digitales normalizados, auténticos y estables. Debido a la rápida obsolescencia de equipos y programas informáticos, y de la incertidumbre en torno a los recursos, responsabilidad y métodos para su conservación, así como una falta de legislación que ampare estos procesos que pueden hacer que desaparezca el patrimonio digital, se pretende adoptar unas medidas jurídicas comunes mundialmente, económicas y técnicas para salvaguardarlo.

ESTÁNDARES Y DIRECTRICES

Recopilación de estándares y directrices que sirven de guía de buenas prácticas para la puesta en marcha de planes de preservación.

Administrative Metadata for Digital Still Images / Harvard University Library (2002-2004).

<http://preserve.harvard.edu/resources/imagemetadata.pdf>

Su finalidad es establecer metadatos para preservar imágenes fijas digitales. Para ello, ImageMetadata proporciona descripciones de elementos; facilita la gestión y conservación de imágenes en DRS (Digital Repository Service) con especial interés en conservar la calidad de las mismas (imagen, color, tamaño) y establece los datos que deben describirse para que sean procesadas y mantener su preservación.

Data Dictionary - Technical Metadata for Digital Still Images / NISO (2006)
<http://www.niso.org/kst/reports/standards/>.

Es un diccionario técnico que establece metadatos, para ser utilizados en la conservación de la imagen digital de calidad. Ha sido diseñado para fijar los estándares a aplicar y para facilitar la interoperabilidad entre distintos sistemas, servicios y softwares, para desarrollar el acceso a colecciones de imágenes digitales de calidad, conservarlas y permitir su uso.

DRS Digital Repository Services

Digital Repository Services User Manual for Data Loading / Harvard University Library (revisado en 2009).

http://hul.harvard.edu/ois/systems/drs/load_manual/drs_load_manual.pdf

Es el manual propio de la Universidad de Harvard para introducir datos de imágenes digitales en la versión 5.3, en el que explica qué datos han de introducirse en cada caso para la identificación de las imágenes, su preservación y consulta, conforme a las características que tenga cada una. Estas imágenes que se introduzcan pueden ser digitales directamente o ser cambiadas de formato analógico a digital. El archivo donde son insertadas es XML (batch.xml). También son enlazados los objetos unos con otros para expresar las relaciones entre ellos. Cada objeto tiene varios "batch" que lo describen. Objetivos: Describir bien cada imagen, conforme a los metadatos existentes para su consulta y conservación a lo largo del tiempo.

DRAMBORA (2008)

Digital Repository Audit Method Based on Risk Assessment
<http://www.repositoryaudit.eu/>

Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) es una herramienta que ha lanzado digitalPreservationEurope (DPE) y Digital Curation Center (DCC) para analizar un depósito digital, determinando sus capacidades, debilidades y fortalezas. Puede descargarse gratuitamente junto con los tutoriales de la aplicación, que consiste en establecer una metodología para facilitar un sistema de auditoría y certificación interna de depósitos digitales, basada en análisis de riesgos.

Fad Guidelines Initiative (2009)

<http://www.digitizationguidelines.gov/>

Es un proyecto cooperativo de agencias federales de Estados Unidos (coordinado por la Library of Congress), que se constituyen en grupo en 2007, para digitalizar documentos con contenido histórico. Consiste en establecer líneas comunes de actuación, así como fijar una metodología para digitalizar contenido histórico.

ISO (OAIS) (2002)

<http://public.ccsds.org/publications/archive/650x0b1.pdf>

Open Archival Information Systems (OAIS) dicta pautas para la conservación y almacenamiento perpetuo de objetos digitales relevantes, así como para su gestión.

METS (2007)

Metadata Encoding & Transimission Standard / Library of Congress

<http://www.loc.gov/standards/mets/>

http://www.loc.gov/standards/mets/METSOverview_spa.html (en español).

Metadata Encoding & Transmission Standard es un formato estandarizado para transmisión de metadatos, pensado para el envío de ficheros, imágenes y objetos multimedia en una biblioteca digital que utiliza la estructura XML. Está desarrollado por Network Development and MARC Standards Office de la Library of Congress. Consiste en la codificación de metadatos necesarios para la gestión de objetos digitales y su intercambio entre depósitos digitales (o entre los depósitos y sus usuarios). Sin metadatos estructurales, las imágenes y los archivos de texto que conforman el objeto digital tienen poca utilidad, y sin los metadatos técnicos relativos al proceso de digitalización, los usuarios no pueden evaluar en qué medida la obra digital es un fiel reflejo del original impreso.

MODS (2002)

Versión 3.3 (2009)

Metadata Object Description Schema / Library of Congress

<http://www.loc.gov/standards/mods/>

Metadata Object Description Schema propone los elementos bibliográficos necesarios para describir y conservar un objeto digital en bibliotecas electrónicas. Está mantenido por Network Development and MARC Standards Office de la Library of Congress. Es otra opción similar a Dublin Core, aunque más sofisticada, para gestionar también metadatos descriptivos. Su esquema contempla el uso de datos seleccionados de registros existentes en MARC y

permite la creación de nuevos registros de descripción de recursos. Incluye algunos campos de MARC y utiliza nombres de etiquetas en lugar de etiquetas numéricas, reagrupando elementos del formato bibliográfico MARC 21. Las autoridades van en MADS (Metadata Authority Description Standard) y el lenguaje de su arquitectura es XML.

NARA Technical guidelines for digitizing archival materials for electronic access (2004)

<http://www.archives.gov/preservation/technical/guidelines.pdf>

Las directrices técnicas de NARA (National Archives and Records Administration de Estados Unidos) para la digitalización de materiales de archivo para el acceso electrónico (NARA technical guidelines for digitizing archival materials for electronic access) fueron redactadas en 2004 por Steven Puglia, Jeffrey Reed, y Erin Rhodes.

National Library of New Zealand Preservation Metadata Schema

<http://www.natlib.govt.nz/catalogues/library-Documents/preservationmetadata-revised>

Publicado en junio de 2003 por la National Library of New Zealand, este esquema de metadatos de preservación detalla los elementos de datos necesarios para sustentar la preservación de objetos digitales y constituir los pilares para el diseño de la base de datos del repositorio y de los sistemas para recoger y almacenar los metadatos de preservación. Incorpora una serie de elementos de datos necesarios para gestionar los metadatos además de incorporar los metadatos relativos al objeto digital en sí mismo. El objetivo ha sido elaborar un documento técnico que sirva como plantilla de implementación, manteniendo a la vez la coherencia con las normas que se están desarrollando a nivel internacional en torno a los metadatos de preservación.

PDF: <http://www.natlib.govt.nz/downloads/metaschema-revised.pdf>

OAI-PMH

The Open Archives Initiative Protocol for Metadata Harvesting
[Protocol Version 2.0]

<http://www.openarchives.org/OAI/openarchivesprotocol.html>

OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) es un protocolo desarrollado por la Open Archives Initiative. Se usa para recolectar las descripciones de metadatos de los registros en un archivo de forma que los servicios puedan ser construidos usando metadatos procedentes de muchos archivos.

PREMIS

Data Dictionary version 2.0

<http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>

<http://www.mcu.es/bibliotecas/MC/PREMIS/Presentacion.html>

PREMIS significa "PREservation Metadata: Implementation Strategies" que es el nombre de un grupo de trabajo internacional patrocinado por OCLC y RLG desde 2003-2005. Ese grupo de trabajo elaboró un informe titulado PREMIS Data Dictionary for Preservation Metadata que incluye un diccionario de datos y una explicación sobre los metadatos de preservación. En marzo de 2008, se publicó una segunda versión actualizada.

Preservation Management of Digital Materials

Preservation Management of Digital Materials: A Handbook / DPC (Digital Preservation Coalition)

<http://www.dpconline.org/graphics/handbook/>

La Digital Preservation Coalition se estableció en 2001 para promover acciones conjuntas dirigidas a realizar cambios urgentes para asegurar la preservación digital de los recursos en el Reino Unido y, para trabajar con otros en el plano internacional con el fin de asegurar la memoria digital y el conocimiento.

Recommendation For Space Data System Standard "Reference Model for an Open Archival Information System" CCSDS 650.0B-1 Blue Book

<http://public.ccsds.org/publications/archive/650x0b1.pdf>

El modelo OAIS (Open Archival Information System) está elaborado por el Consultative Committee for Space Data Systems de la NASA. Posteriormente se ha convertido en el estándar ISO 14721:2003 y es el modelo que están adoptando diferentes instituciones que tienen la misión explícita de preservar la información de cualquier tipo.

Safeguarding Australia's web resources: guidelines for creators and publishers / National Library of Australia

<http://www.nla.gov.au/guidelines/webresources.html>.

Las directrices para salvaguardar los recursos web de Australia elaboradas por la Biblioteca Nacional de Australia en 2002, están destinadas principalmente a los creadores y editores que no tienen bien establecido el sistema de gestión de datos digitales.

TRAC

Trustworthy Repositories Audit & Certification: Criteria and Checklist

<http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>

Criterios para la certificación de repositorios elaborados por CRL, OCLC y los National Archives and Records Administration de USA. Están basados, entre otras iniciativas, en el grupo de trabajo NESTOR. Tienen unos objetivos (preservación a largo término de objetos digitales) y una estructura muy similar a NESTOR por lo que pueden ser complementarios. Los criterios de certificación se dividen en: Infraestructura de la organización, Gestión de objetos digitales, Tecnologías, infraestructura técnica y seguridad.

Web at Risk collection planning guidelines

Web at Risk collection planning guidelines / Kathkeen R. Murray; Inga K. Hiesh (2006).

http://www.digitalpreservation.gov/partners/web-at-risk/high/cpg_final_31may2006.pdf

El Web-at-Risk Project es un depósito digital diseñado para preservar a perpetuidad el patrimonio cultural de Estados Unidos, que está promovido por la Biblioteca Digital de California (CDL), aunque participan muchas bibliotecas e investigadores en el proyecto. Es un depósito digital, legalmente aprobado y permitido, que dispone de WAS (Web Archiving Service), que es una aplicación web diseñada para capturar y preservar el contenido diverso que interese y organizarlo, citando su fuente principal.

PROYECTOS

Recopilación de los proyectos que se están llevando a cabo en el panorama mundial, algunos de los proyectos referenciados ya han finalizado pero de alguna manera siguen estando vigentes por lo que es interesante conocerlos.

CAMiLEON

Creative Archiving at Michigan and Leeds Emulating the Old On the New

<http://www.si.umich.edu/CAMiLEON>

El proyecto CAMiLEON (Creative Archiving at Michigan & Leeds: Emulating the Old on the New) desarrolla y evalúa una serie de estrategias técnicas para la preservación a largo plazo de materiales digitales. A raíz de un proyecto anterior, el proyecto CEDARS, la Universidad de Leeds y la Universidad de Michigan, desarrollaron el proyecto CAMiLEON con el objetivo de investigar la emulación como estrategia de preservación digital. El proyecto fue patrocinado por JISC (Joint Information Systems Committee) y NSF (National Science Foundation).

CASPAR

Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval

<http://www.casparpreserves.eu/>

CASPAR es un proyecto de investigación financiado por la Unión Europea con el objetivo de estudiar, implementar y difundir soluciones de preservación digital basadas en el modelo de referencia OAIS.

CEDARS

Curl exemplars in digital archives

<http://www.leeds.ac.uk/cedars>

El proyecto Cedars comenzó en abril de 1998 y finalizó en Marzo de 2002. Fue desarrollado por las universidades de Leeds, Cambridge y Oxford. Su principal objetivo fue investigar en temas de preservación digital dentro del ámbito de la adquisición de objetos digitales, su permanencia a largo plazo, la descripción, y el acceso eventual. El informe del grupo de trabajo del proyecto es una investigación dirigida a estudiar los casos de la gran variedad de materiales digitales depositados y, a establecer recomendaciones básicas para el uso de la migración y la emulación y las estrategias de preservación. El informe también aporta una guía y recomendaciones sobre temas de organización y gestión. A raíz de este proyecto, la Universidad de Leeds y la Universidad de Michigan, desarrollaron el proyecto CAMiLEON con el objetivo de investigar la emulación como estrategia de preservación digital.

CLOCKSS (Controlled Lots Of Copies Keep Stuff Safe): 2006-2007.

http://edina.ac.uk/projects/clockss_summary.html

CLOCKSS es una alianza de bibliotecas y editores formada en 2005 para asegurar el acceso a largo plazo del trabajo académico en formato digital. Al moverse cada vez más el contenido online, creció la preocupación de que este contenido digital no estuviera siempre disponible. El Proyecto CLOCKSS se creó para abordar este problema investigando sobre cómo crear un archivo sin fines lucrativos, dirigido por la comunidad, seguro, con contenido publicado en la web que podría aprovecharse cuando el acceso a contenidos de interés

académico desaparecieran por razones de desastres naturales, tecnológicos o por cualquier fallo humano.

DCC (Digital Curation Centre)

<http://www.dcc.ac.uk>

Digital Curation Centre (DCC) apoya a las instituciones del Reino Unido que almacenan, gestionan y preservan datos digitales tales como el registro científico y el patrimonio documental creados en formato digital. DCC trabaja con otros profesionales para asegurar la mejora continua y el uso a largo plazo de los datos digitales. La conservación digital consiste en mantener y añadir valor a una parte importante de la información digital para el uso actual y futuro; en concreto, en lo que se refiere a la gestión activa y a la evaluación de los datos sobre el ciclo de vida de las materias científicas y académicas. DCC es el centro por excelencia en conservación y preservación digital en el Reino Unido. Es una fuente autorizada de asesoramiento de expertos para la comunidad. Es también un proveedor de variados recursos, software, herramientas y servicios de apoyo. Empezó su actividad con una primera fase de objetivos en noviembre de 2003.

DELOS (Network of Excellence on Digital Libraries)

<http://www.delos.info>

DELOS es una red de bibliotecas digitales, patrocinada parcialmente por la Comisión Europea en el marco del Programa de Tecnologías de la Sociedad de la Información (IST). El proyecto comenzó en enero de 2005. Los principales objetivos de DELOS son la investigación, cuyos resultados son de dominio público, y la transferencia de tecnología, a través de acuerdos de cooperación con las partes interesadas. Que todos los ciudadanos, en cualquier lugar y en cualquier momento, deberían tener acceso a Internet conectados a dispositivos digitales para buscar todo el conocimiento humano, independientemente de las barreras de tiempo, lugar, cultura o idioma ha sido una visión de DELOS desde su creación. DELOS se organiza en grupos de trabajo, uno de ellos es *Preservation* (WP6). Su objetivo es proporcionar el marco metodológico y la teoría de garantizar que las bibliotecas digitales de investigación aborden aspectos de preservación y que las bibliotecas digitales incorporen en sus diseño elementos de conservación.

DLF (Digital Library Federation)

<http://www.diglib.org>

DLF es un programa del Council on Library and Information Resources (CLIR) de los Estados Unidos, que tiene por objetivo reunir expertos de las comunidades de científicos, bibliotecarios e informáticos para abordar una serie de iniciativas como son: uso, preservación, normalización y arquitecturas de bibliotecas digitales; archivo de revistas electrónicas; colecciones en línea para su utilización en la docencia; servicios de Internet que van más allá del acceso a los recursos por parte de los científicos; estudio de los papeles que jugarán en el futuro los bibliotecarios; etc.

DPE (Digital Preservation Europe)

<http://www.digitalpreservationeurope.eu>

Digital Preservation Europe (DPE) es un proyecto financiado por la Unión Europea que busca crear colaboración y sinergias entre una variedad de

iniciativas nacionales existentes en el área de investigación europea. Trata de solucionar las necesidades de mejora en la coordinación, cooperación y consistencia en las actividades que se llevan a cabo para asegurar una preservación eficaz de los materiales digitales. En concreto su trabajo está encaminado a definir el perfil de preservación digital, promover actuaciones globales en materia de valores añadidos de la preservación en los estados miembros, facilitar la formación profesional, facilitar el intercambio de resultados de investigación, desarrollar y promover una agenda de investigación a nivel europeo y ayudar tanto a ciudadanos como a especialistas a reconocer el papel central que la preservación digital juega en nuestras vidas. Entre sus productos tenemos DRAMBORA, una metodología destinada a facilitar un sistema de auditoría y certificación interna de repositorios de documentos digitales basada en el análisis de riesgos; PLATTER (Planning Tool for Trusted Electronic Repositories) que proporciona los fundamentos para que un repositorio digital planee el desarrollo de sus objetivos; ErpaNet un depósito de eprints especializado en preservación.

Economics of digital preservation: a project of OCLC Research

<http://www.oclc.org/research/projects/digipres/economics.htm>

Este proyecto de OCLC investiga aspectos relacionados con la economía de la preservación digital. Su resultado se ha concretado en el libro blanco: Lavoie, Brian F. 2003. *The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making*. Disponible en:

<http://www.oclc.org/research/projects/digipres/incentives-dp.pdf>

(PDF:746.8K/51pp.). También en la publicación relacionada: Lavoie, Brian F. 2004. "Of Mice and Memory: Economically Sustainable Preservation for the Twenty-first Century." En *Access in the Future Tense*. Washington, D.C.: Council on Library and Information Resources. Disponible en: <http://www.clir.org/pubs/abstract/pub126abst.html>. El proyecto parece estar finalizado en estos momentos.

IFLA. Preservation and Conservation Section

<http://www.ifla.org/VII/s19/index.htm>

Esta sección de la IFLA está relacionada con la preservación del patrimonio documental mundial. Proporciona un forum internacional para que cualquier biblioteca intercambie, desarrolle y difunda sus conocimientos y experiencias relativas a teorías, políticas y prácticas para la preservación del conocimiento registrado independientemente de los medios materiales utilizados. En el último plan estratégico correspondiente a 2009 se recogen seis objetivos con sus respectivas acciones asociadas. Entre ellos: promover entre las bibliotecas, archivos, museos y el público en general, un mejor conocimiento y la importancia de la preservación y conservación del patrimonio; promover la educación y el desarrollo profesional en el campo de la preservación; promover buenas prácticas y normalización en las actividades de preservación; potenciar y facilitar el establecimiento de estrategias y programas nacionales de preservación; presionar para el desarrollo de soluciones y métodos para la preservación a largo plazo de materiales bibliotecarios que se encuentran en formato digital. Entre sus publicaciones destaca *Guidelines for Digitization Projects for collections and holdings in the public domain, particularly those held by libraries and archives*, que ha sido traducido al castellano y al catalán. Además organiza congresos y publica un Informe Anual.

IIPC (International Internet Preservation Consortium)

<http://netpreserve.org/about/index.php>

La misión de IIPC, como consorcio de bibliotecas nacionales, es adquirir, preservar y hacer accesible el conocimiento e información de Internet para futuras generaciones, promocionando los intercambios a nivel global y las relaciones internacionales. Para cumplir con esta misión está trabajando en los siguientes objetivos: coleccionar una gran muestra de contenidos de Internet de todo el mundo para ser conservados en una forma que permita su archivo, garantice su seguridad y permita su acceso a lo largo del tiempo; promover el desarrollo y uso de herramientas comunes, técnicas y estándares que permitan la creación de archivos internacionales; dar soporte a bibliotecas nacionales para llevar a cabo el archivo y preservación de contenidos distribuidos en Internet.

Kopal: Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen

<http://kopal.langzeitarchivierung.de/index.php.en>

Se trata de un proyecto llevado a cabo en el período 2004-2007 con el objetivo de desarrollar una solución tecnológica y organizacional para asegurar la disponibilidad a largo plazo de publicaciones electrónicas. Para ello la integración de forma transparente en los sistemas de bibliotecas existentes y su reutilización a través de instituciones de la memoria juegan un papel crítico. En la implementación del sistema se adoptaron estándares internacionales sobre metadatos y preservación a largo plazo. Los materiales digitales que sirvieron de campo de pruebas para el proyecto procedían de la German National Library y de Niedersaechsische Staats und Universitaetsbibliothek Goettingen. Entre los resultados obtenidos se encuentra el paquete de software koLibRI (kopal Library for Retrieval and Ingest), una biblioteca de herramientas escrita en Java para interactuar con el sistema DIAS de IMB. La intención es que otras instituciones interesadas en la preservación de sus contenidos digitales puedan adaptarlo a sus sistemas y arquitecturas.

Memory of the World - Preserving Our Digital Heritage / UNESCO

http://portal.unesco.org/ci/en/ev.phpURL_ID=1538&URL_DO=DO_TOPIC&URL_SECTION=201.html

Memory of the World – Preserving our Digital Heritage es el programa de la UNESCO que tiene como objetivo la preservación y difusión de valiosos fondos archivísticos y colecciones de bibliotecas de todo el mundo. La misión de este programa se puede resumir en los siguientes puntos: facilitar la preservación, utilizando las técnicas más apropiadas, del patrimonio documental mundial, bien a través de una asistencia directa, de asesoramiento o de la búsqueda de sponsors; colaborar en conseguir un acceso universal al patrimonio documental a través de fomentar la digitalización y puesta en Internet de contenidos, la publicación y distribución de documentos, el estudio de las limitaciones legales para el acceso a los archivos, etc; Incrementar la concienciación mundial sobre la existencia y significado del patrimonio documental, etc. Entre los proyectos creados bajo los auspicios de este programa podemos citar la *Memoria de Iberoamérica*. Es un proyecto de bibliotecas nacionales de doce países para realizar un inventario de cerca de seis mil periódicos del siglo XIX así como su digitalización y conservación.

NDIIPP: National Digital Information Infrastructure and Preservation Program
<http://www.digitalpreservation.gov>

En diciembre de 2000 el Congreso de los Estados Unidos pidió a la Library of Congress que dirigiese el NDIIPP. La misión del National Digital Information Infrastructure and Preservation Program de la Biblioteca del Congreso de Estados Unidos es desarrollar una estrategia a nivel nacional para coleccionar, preservar y hacer accesibles contenidos digitales, especialmente información que ha sido creada únicamente en formato electrónico, para las generaciones actuales y futuras. Para conseguirlo recurre a un conjunto de socios tanto privados como públicos con los cuales desarrollar una serie de iniciativas concretas como: Archivo digital y preservación a largo término (DIGARCH) uno de cuyos proyectos es Digital Engineering Archives, arquitecturas técnicas como JHOVE2: A Next Generation Architecture for Format Aware Digital Preservation Processing; Políticas de preservación digital dentro de la que se ha publicado el estudio: International Study on the Impact of Copyright Law on Digital Preservation al que es posible acceder y descargar a través del siguiente enlace

http://www.digitalpreservation.gov/library/resources/pubs/docs/digital_preservation_on_final_report2008.pdf ; Conservación de obras creativas como por ejemplo Preserving Digital Cartoons y finalmente, conservación de información gubernamental del estado.

NESTOR (Network of Expertise in Long-Term Storage of Digital Resources)

<http://www.langzeitarchivierung.de/index.php?newlang=eng>

Nestor es un proyecto alemán creado con la misión de asegurar la preservación de los recursos digitales en Alemania y trabajar con otras instituciones a nivel internacional para garantizar el acceso a la memoria digital mundial. Entre sus objetivos concretos está el servir de punto de unión y comunicación a todas las partes que trabajan en Alemania sobre preservación a largo término de los documentos electrónicos. Entre sus resultados está un portal temático en colaboración con PADI donde se recogen más de 2000 publicaciones sobre preservación. Su catálogo para la creación de repositorios fiables, identifica criterios que facilitan la evaluación de repositorios y se basa en la fiabilidad de los contenidos y del propio repositorio tanto a nivel organizativo como técnico. Su objetivo es asegurar la preservación de los contenidos, almacenados como objetos digitales, a largo plazo. Considera un repositorio como aquella institución que ha asumido la responsabilidad de la preservación y accesibilidad a largo plazo de objetos digitales asegurando que puedan ser utilizados en el futuro por una comunidad de usuarios o grupo específico. Los criterios de evaluación que se plantean se articulan en tres apartados: marco organizativo, gestión de objetos e infraestructura y seguridad.

PADI: Preserving Access to Digital Information / National Library of Australia

<http://www.nla.gov.au/padi/>

PADI (Preserving Access to Digital Information) es un proyecto de la National Library of Australia cuyo objetivo es proporcionar mecanismos que ayuden a asegurar que la información en formato digital sea gestionada con la consideración adecuada para su preservación y futuro acceso. Para ello trabaja en las siguientes direcciones: facilitar el desarrollo de estrategias y guías para la preservación y acceso a la información digital, desarrollar y mantener un sitio web para propósitos de información y promoción; proporcionar un forum para la

cooperación intersectorial sobre actividades relativas a la promoción del acceso a la información digital a largo plazo. De esta forma la página web de PADI es una pasarela temática donde se recogen multitud de recursos sobre preservación y conservación. Se presentan clasificados en categorías tales como: Estrategias de preservación digital, metadatos, identificadores persistentes, gestión de derechos de propiedad intelectual, formatos y medios, etc. Los recursos pueden ser remitidos por cualquier persona mediante un formulario si bien deben cumplir con los requisitos especificados en una guía de selección.

PADICAT. Patrimoni Digital de Catalunya. Biblioteca de Catalunya
<http://www.padicat.cat/es/index.php>

La Biblioteca de Catalunya como responsable de compilar, procesar y difundir el patrimonio bibliográfico de Cataluña inició este proyecto en 2005. El proyecto consiste en capturar, procesar y dar acceso permanente a toda la producción cultural, científica y de carácter general catalana producida en formato digital. En definitiva, el objetivo de PADICAT es archivar la web catalana.

PLANETS (Preservation and Long-term Access through Networked Services)
<http://www.planets-project.eu/>

Planets (Preservation and Long-term Access through Networked Services) es un proyecto iniciado en 2006 por un período de cuatro años y financiado por la Unión Europea dentro del VI Programa Marco con la misión de tratar los retos que tiene planteados la preservación digital. El primer objetivo para Planets es construir servicios prácticos y herramientas para ayudar a asegurar el acceso a largo plazo a nuestro patrimonio cultural y científico que existe en formato electrónico. Según su presentación, creará un marco sostenible para permitir la preservación de los contenidos digitales, aumentando la capacidad europea para asegurar el acceso en perpetuidad a su información digital. En concreto los resultados del proyecto serán: servicios de planificación de la preservación que permitan a las organizaciones definir, evaluar y llevar a cabo la preservación; metodologías, herramientas y servicios para la caracterización de los objetos digitales; un marco de interoperabilidad para integrar de forma transparente herramientas y servicios en un sistema distribuido; un campo de pruebas para proporcionar unas evidencias consistentes y coherentes para la evaluación objetiva de diferentes protocolos y planes de preservación; un programa de difusión y diseminación del proyecto que permita la mayor adopción de resultados posible entre la comunidad de usuarios. En conclusión el proyecto permitirá a las organizaciones mejorar el proceso de toma de decisiones relativas a la preservación, asegurar el acceso a largo plazo a sus contenidos digitales y controlar los costes de las acciones de conservación a través de una progresiva automatización y una infraestructura escalable.

PLAN DE PRESERVACIÓN DIGITAL PARA GREDOS

GREDOS

GREDOS es el acrónimo de Gestión del Repositorio Documental de Salamanca (de la Universidad de Salamanca). Se trata de una plataforma de gestión del conocimiento de esta institución y el portal de acceso a información patrimonial, docente, científica e institucional, en formato digital.

Difunde ofreciendo la consulta en línea de documentos digitales con contenidos históricos, científicos, didácticos e institucionales. Los repositorios son la herramienta de la que se sirven las instituciones académicas y científicas para poner documentación en libre acceso. La Universidad de Salamanca firmó la Declaración de Berlín sobre el Acceso Abierto al Conocimiento en Ciencias y Humanidades, lo que implica un compromiso para ofrecer las condiciones tecnológicas y de apoyo institucional para que la producción científica y documental generada en esta universidad pueda ser difundida libremente. El sistema de Gestión del Repositorio Documental de la Universidad de Salamanca Gredos tiene carácter integrador, ya que reúne en un mismo repositorio toda la producción digital de la Universidad de Salamanca. Está estructurado en cuatro secciones:

1. Biblioteca digital, con contenidos históricos y patrimoniales: manuscritos, libros antiguos, prensa y revistas digitalizadas.
2. Repositorio científico, donde se alberga la producción científica de la Universidad de Salamanca: tesis doctorales, artículos, libros electrónicos o actas de congresos.
3. Repositorio docente, con recursos de carácter didáctico: materiales educativos, asignaturas abiertas, audiovisuales, tutoriales. En esa sección se incluyen también contenidos de la Fundación Germán Sánchez Ruipérez.
4. Archivo institucional, de carácter administrativo: normativas universitarias, archivos fotográficos, expedientes, guías académicas, documentos históricos y archivos personales de investigadores de la Universidad de Salamanca.



Figura 1. Página de inicio con las cuatro grandes comunidades de Gredos

Se ha partido de la intención de que cualquier documento generado en la Universidad, que se quiera compartir pueda estar disponible en el repositorio con los requisitos mínimos impuestos por la licencia seleccionada. Todos los documentos depositados en GREDOS están protegidos con una licencia Creative Commons.

Como podemos observar se habla siempre de difusión y consulta de la documentación digital. Gredos almacena y difunde en la actualidad más de 60.000 documentos digitales de la Universidad de Salamanca, ¿podemos asegurar que estos documentos perdurarán y serán accesibles a largo plazo?

Hoy por hoy no lo podemos asegurar, pero será necesario comenzar a implementar un plan de preservación para el repositorio que asegure que la documentación albergada en él, que es fruto del trabajo de toda la comunidad universitaria y reflejo de su investigación y docencia, no se pierda en el tiempo siendo irrecuperable para las generaciones venideras.

Los repositorios institucionales que albergan toda la producción digital de la institución a la que representan, constituyen además de un medio de difusión en acceso abierto una herramienta de preservación de la documentación allí depositada. Si se consigue elaborar un plan eficaz que contenga y aplique los protocolos, estándares y directrices existentes, que esté de acuerdo a las políticas de la institución y que asegure su continuidad en el tiempo, se conseguirá llegar a poner en marcha un plan de preservación digital a largo plazo.

Volvemos a repetir aquí los problemas con los que nos vamos a encontrar, que suelen ser comunes a este tipo de proyectos y que en el caso de la Universidad de Salamanca serán los fundamentales a la hora de acometer e intentar poner en marcha un plan de preservación digital a largo plazo:

- Legales: si no tenemos el permiso del titular no podremos preservar un recurso en Gredos reproduciéndolo o reformateándolo.
- Económicos: si no contamos con los medios necesarios no podremos garantizar la perdurabilidad de los documentos en Gredos a lo largo de los años.
- Institucionales: si no se asegura el compromiso institucional permanente de la Universidad, si no están convencidos todos los implicados que detenta los derechos de la necesidad de colaborar, si no podemos verificar si se cumple todo lo que se promete, no podremos preservar a largo plazo el contenido de Gredos.

En cuanto a los problemas técnicos con los que nos podemos encontrar a la hora de preservar la información, estarán derivados en gran parte por la deficiente parametrización y aplicación del software que estemos utilizando. Es por ello, que debemos ser extremadamente escrupulosos en la aplicación de normas, protocolos y metadatos de preservación que nos ofrezca el sistema que estemos utilizando.

Actualmente, el software de gestión de Gredos es la plataforma DSpace versión 1.5.2. Ante la elaboración de un proyecto de preservación digital a largo plazo debemos conocer a fondo si este software que se está utilizando en la

gestión del repositorio digital se adecúa a las normas y protocolos de preservación básicos para desarrollar dicho plan.

El Ministerio de Cultura Español marca una serie de requisitos en el tema de preservación a la hora de conceder ayudas económicas para la creación de recursos digitales en acceso abierto. En las distintas convocatorias de estas ayudas el Ministerio solicita *“datos específicos de creación de recursos digitales para el establecimiento de datos de preservación conforme al “Diccionario de datos PREMIS” y a la normativa OAIS / ISO 14721”*.

En el desarrollo de este trabajo se va a analizar si DSpace aplica los protocolos y las directrices recomendadas para la preservación: modelo OAIS, METS, y metadatos de preservación (PREMIS).

También se analizará el hardware utilizado para la gestión del repositorio institucional, las medidas de seguridad empleadas y las medidas que serían recomendables utilizar en un sistema de preservación a largo plazo.

PRINCIPIOS DE UN PLAN DE PRESERVACIÓN

Es fundamental establecer unos principios básicos para afrontar un plan de preservación digital, o mejor dicho un compromiso de los gestores y financiadores de la institución para preservar a lo largo del tiempo el repositorio institucional. Se consideran como principios básicos para implementar un plan de preservación en Gredos:

1. Escribir e implementar una política de preservación por cada colección digital, en este caso una política por cada una de las cuatro grandes comunidades de Gredos: Archivo institucional, Biblioteca digital, Repositorio científico y Repositorio institucional.
2. Definir los procedimientos de gestión de cada colección y su contenido.
3. Preservar el contenido en una infraestructura apropiada de preservación digital.
4. Identificar qué colecciones serán mantenidas a lo largo del tiempo.
5. Reconocer que la financiación para la digitalización implica la aceptación de un compromiso institucional que sobreviva al período de financiación.
6. Ser explícito sobre los periodos de retención para los diferentes tipos de contenido digital.

Por su parte la institución financiadora del plan, en este caso la Universidad de Salamanca, debería:

7. Establecer los servicios de conservación del contenido y el uso de métodos apropiados para asegurar que las condiciones se cumplen.
8. Articular con la mayor claridad posible y hacer cumplir los requisitos para la retención de los contenidos.
9. Regular con claridad la obligación de preservar a largo plazo.

En este punto se habla de colecciones digitales del repositorio porque, en último término, el contenido de Gredos se encuentra estructurado en colecciones dentro de cada una de las grandes comunidades que lo forman.

Cada una de estas colecciones debería crearse de acuerdo a criterios sólidos y no caprichosos, debidamente documentados.

A continuación se hace hincapié en los seis principios básicos que atañen a los gestores y financiadores de la institución y que tendrían que desarrollarse en un plan de preservación del repositorio institucional.

Escribir e implementar una política de preservación por cada colección digital de Gredos porque sin una política de preservación escrita la usabilidad a largo plazo, la autenticidad y la accesibilidad de este repositorio digital está en peligro. Cada colección digital debería tener una política para identificar los procedimientos y las partes interesadas. Esta política debe abordar:

- qué contenido se va a preservar y por quién,
- el objetivo de la preservación,
- quién es el responsable de la preservación de los archivos y metadatos relacionados,
- fuentes de financiación para la preservación que se vaya a llevar a cabo,
- cómo deberá medirse el éxito de la preservación.

Una vez aprobado el plan debería ponerse en práctica.

Definir los procedimientos de gestión de cada colección y su contenido porque de no recogerse los procedimientos de gestión de las colecciones definidas y de los contenidos, la usabilidad a largo plazo, la autenticidad y la accesibilidad de la colección digital se encuentra en riesgo. El procedimiento de gestión de las colecciones de Gredos y su contenido debería estar documentado en las siguientes áreas:

Creación de la colección inicial de Gredos. Los procedimientos para crear las colecciones digitales iniciales deberían estar bien documentados. Sería preferible un flujo de trabajo automatizado, aunque es aceptable también un procedimiento de la gestión de contenido manual, siempre que esté bien documentado.

Mantenimiento de las colecciones en curso de Gredos. Las colecciones del repositorio requieren un mantenimiento continuo de conservación (por ejemplo, chequear regularmente si el contenido se ha corrompido o se ha perdido). Además se requiere un mantenimiento continuo del contenido intelectual (por ejemplo, la corrección de metadatos descriptivos inexactos). La persona que se encargue del mantenimiento de la preservación probablemente no tendrá el conocimiento ni el dominio necesario para mantener el contenido intelectual de la colección. Los procedimientos y el personal responsable del mantenimiento de la preservación y del mantenimiento de los contenidos deberían estar documentados.

Preservar el contenido del repositorio en una infraestructura apropiada de preservación digital porque de no mantener las colecciones digitales en una infraestructura apropiada la usabilidad y accesibilidad del repositorio digital está en riesgo. La preservación digital a largo plazo de Gredos no se va a lograr únicamente a través de las copias de seguridad, la redundancia del sistema de acceso o la replicación byte. Aunque estas acciones puedan ofrecer cierta

protección a corto y medio plazo y a menudo sean elementos de preservación a largo plazo, no son elementos suficientes (ver figura 2).

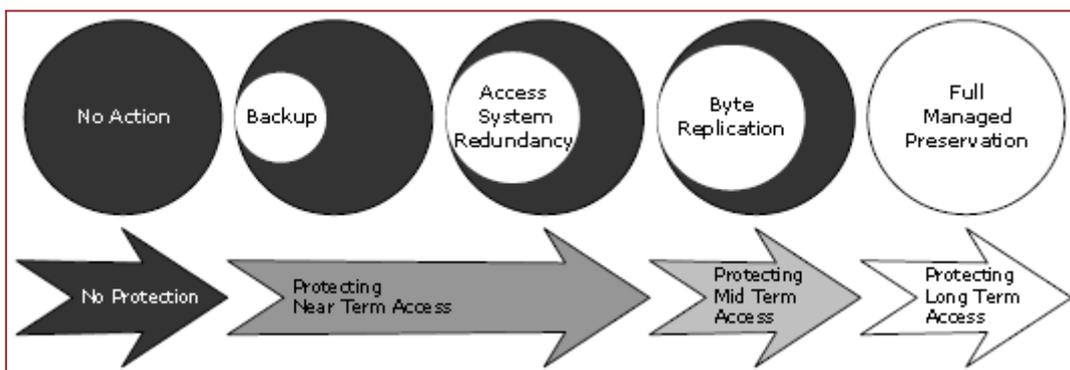


Figura 2. Tomado de JISC Digitisation Programme: Preservation Study April 2009

Si el repositorio institucional no pudiera soportar la gestión por su cuenta de un sistema de preservación digital a largo plazo, se debería considerar la posibilidad de construir una infraestructura de preservación digital dentro de la institución, en este caso la Universidad de Salamanca, o entre varias organizaciones, por ejemplo BUCLE (Bibliotecas Universitarias de Castilla y León), o la utilización de servicios ofrecidos por terceros, por ejemplo DRAMBORA.

Identificar qué colecciones serán mantenidas a lo largo del tiempo. Sin un plan de sostenibilidad, la usabilidad y la accesibilidad a largo plazo se encuentran en riesgo. Aunque mantener el acceso público no es un requisito previo para mantener el contenido principal, sí es un elemento efectivo para hacerlo. Como se indica en las recomendaciones anteriores, las colecciones digitales requieren mantenimiento continuo para que puedan preservarse a largo plazo para la comunidad designada. El tiempo durante el que una colección debe conservarse tiene que estar definido al inicio del proyecto, de forma que se puedan presupuestar los fondos necesarios y que la financiación se identifique y documente como un elemento en el plan de preservación del repositorio. Lo más importante, es que la institución reconozca que para que las colecciones digitales del repositorio institucional sean conservadas se requiere mantenimiento continuo y un compromiso financiero para apoyar este mantenimiento que deberá estar asegurado de forma explícita.

Ser explícito sobre los periodos de retención para los diferentes tipos de contenido digital. No todo el contenido creado con la financiación del proyecto justifica la retención a largo plazo. Algunos materiales de enseñanza, docencia o administrativos que están incluidos en el repositorio institucional de la Universidad de Salamanca, pueden tener un ciclo de vida más corto que los materiales contenidos en la Biblioteca digital y en el Repositorio científico, por lo que sería necesario establecer si hay que tenerlos en cuenta en la política de preservación digital a largo plazo o bien establecer un tiempo de retención para este tipo de material digital, o por el contrario mantenerlos dentro del plan de preservación.

FACTORES TECNOLÓGICOS

CREACIÓN DE RECURSOS DIGITALES EN GREDOS

Es claro que para proceder a la creación de recursos digitales que permitan la máxima visibilidad y accesibilidad en la red, así como, el intercambio de información entre repositorios, o la recolección de metadatos en diferentes repositorios o recolectores, es necesario el uso de esquemas de metadatos de diferente tipo. Así, para crear los metadatos descriptivos mínimos del protocolo OAI-PMH sólo es necesario utilizar Dublin Core sin cualificar bajo el esquema “dc_oai”, pero si se pretende transmitir la relación completa de los componentes de un objeto digital y su estructura será necesario contar con el esquema METS.

En el caso de Gredos, se están utilizando metadatos Dublin Core Cualificado, aplicando el protocolo OAI y también se cuenta con el esquema METS. Para la aplicación en Gredos del esquema Dublin Core Cualificado se redactó previamente a su aplicación una guía de buenas prácticas, *Qualified Dublin Core Metadata Best Practices for GREDOS*, con la intención de normalizar el uso de DCQ en la descripción de las distintas colecciones del repositorio (García-Peñalvo et al., 2010).

Asimismo, en la creación de recursos digitales es necesario incluir todo un conjunto de datos administrativos y técnicos definidos en los correspondientes esquemas de metadatos que se ajusten a diferentes funcionalidades, entre ellas las propias de la preservación digital a largo plazo.

Existe una guía de buenas prácticas para la construcción de colecciones digitales de calidad y para la aplicación de esquemas de metadatos a estas colecciones que es *A Framework of Guidance for Building Good Digital Collections* asumidas por la Digital Library Federation, y que también sería recomendable seguir y aplicar en la creación de metadatos del repositorio institucional de la Universidad de Salamanca, además de la guía sobre Dublin Core ya citada anteriormente. En esta publicación se definen seis principios a los que deben ajustarse los metadatos (“good metadata”) utilizados en una colección digital:

1. Deben seguir la normativa propia de una comunidad y ser apropiados para el tratamiento de los materiales, para los usuarios y para el uso actual y futuro de la colección.
2. Deben ser interoperables.
3. Deben disponer de control de autoridades y de normas específicas para la elaboración de contenido tanto de las descripciones, como de los objetos y sus relaciones.
4. Deben incluir una declaración sobre las condiciones de uso de los objetos digitales.
5. Deben soportar la preservación a largo plazo de todos los objetos incluidos en una colección.
6. Deben, en tanto que objetos en sí mismos, tener las características de autoridad, autenticidad, persistencia, y estar identificados unívocamente.

SISTEMA Y PRESERVACIÓN DIGITAL CON DSPACE

Cualquier repositorio es en parte producto del software de gestión que se esté utilizando. Tanto la estructura, los formatos, los registros, hasta la preservación van a estar condicionados por el software que se utilice para gestionarlo. En el caso que nos ocupa se trata del software Dspace.

Dspace usa una interfaz web de fácil navegación basada en una estructura jerárquica para almacenar y recuperar la información en 5 niveles:

- Comunidades: conjunto de colecciones y subcomunidades. Ej: Departamentos, Laboratorios, Centros de Investigación...
- Colecciones: conjunto de ítems.
- Ítems: es el elemento básico. Se compone de bundles o paquetes de archivos.
- Bundle: compuesto por bitstreams o archivos.
- Bitstreams (archivos). Ej: Original, Text, Licence. Bitstreams (información técnica. ej. formato). Ej: PDF, JPEG.

Como se ha venido diciendo en el desarrollo de este trabajo, la preservación digital es el conjunto de estrategias, procesos y técnicas que dan respuesta a los problemas que plantea la conservación de los materiales digitales y de los medios (hardware y software) que se emplean para su almacenamiento y consulta, y que están derivados fundamentalmente de la obsolescencia provocada por la rápida renovación tecnológica y por la inestabilidad de los soportes. Estas técnicas son muy variadas y responden a diferentes situaciones y líneas estratégicas (copias de seguridad, copia de datos en soportes durables, migración, replicación, emulación, etc.), aunque en general están destinadas a mantener los objetos digitales y sus características de acceso a largo plazo.

La preocupación internacional por la preservación digital se concretó en 2002 con la publicación del Reference Model for an Open Archival Information System (OAIS). Se trata de una recomendación para el desarrollo de un amplio consenso sobre los requisitos que debe cumplir un archivo para preservar a largo plazo la información digital y definir la norma ISO Reference Model for an Open Archival Information System (ISO 14721:2003). Un sistema OAIS es un archivo, que consiste en una organización, de equipos humanos y sistemas, que tienen la responsabilidad de preservar información y hacerla disponible para una comunidad específica ("designated community"). Un sistema estará de acuerdo a OAIS si soporta el modelo de información descrito en la norma que no especifica ningún método de implantación. Es importante tener una noción básica del modelo informativo, que nos hará entender con mayor precisión los metadatos necesarios para la preservación en Gredos.

En el modelo OAIS es básico el concepto Paquete de Información (Information Package) que son contenedores conceptuales que agrupan dos tipos de información, la Información de Contenido (Content Information) y la Información de Preservación (Preservation Description Information). La Información de Contenido es el objeto propio de la preservación (y sus procesos) y está formada por los objetos digitales (bits) y la información de representación

asociada (p.e., juego de caracteres) necesaria para que un objeto digital sea comprensible para una comunidad específica (hardware, programas, componentes, etc.).

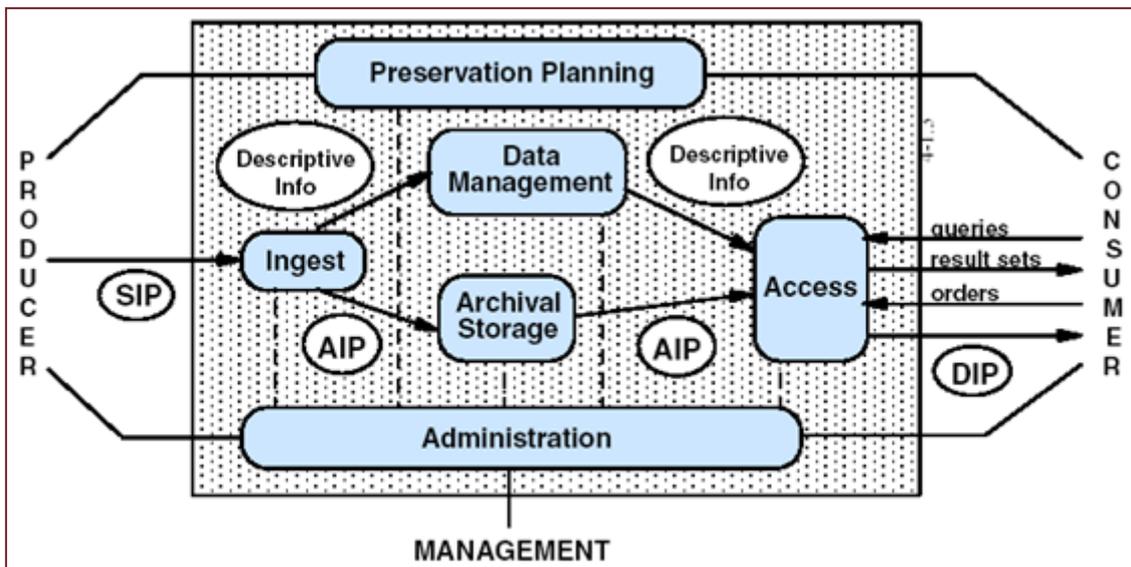


Figura 3. Tomado de Reference Model for an Open Archival Information System (OAIS): CCSDS 650.0-B-1 : Blue Book

La Información de Preservación es la información necesaria para preservar la Información de Contenido y debe proporcionar suficientes datos sobre su procedencia, contexto, referencia y fijeza. La procedencia describe tanto la fuente de la Información de Contenido que la ha custodiado desde su origen, como su historia, incluyendo la historia de los procesos realizados. El contexto describe cómo la Información de Contenido se relaciona con otra información externa al Paquete de Información o con otras Informaciones de Contenido. La Referencia proporciona uno o más identificadores o sistemas de identificadores unívocos, de modo que una determinada Información de Contenido sea perfectamente distinguible de otras. Por último, la fijeza proporciona un envoltorio o escudo protector que protege la Información de Contenido de cualquier alteración o que permite su detección (p.e., mecanismos de suma de verificación). Un sistema OAIS trabaja con paquetes de información que se distinguen según los procesos en los que participan. Así un Submission Information Package (SIP) es un paquete de información que se envía a un sistema OAIS por un productor. Su forma y contenido (información de contenido e información de preservación) deben ser establecidos entre el productor y el sistema OAIS. Una vez que este paquete de información haya sido incorporado al sistema se complementará con las informaciones específicas propias para la gestión de preservación con lo que se transformará en un Archival Information Package (AIP). A su vez un sistema OAIS devolverá a los usuarios que lo soliciten los datos necesarios en forma de un Dissemination Information Package (DIP), de esta forma el usuario podrá distinguir claramente la información solicitada.

Un sistema OAIS debe proporcionar una serie de funciones como son: las funciones de integración (Ingest), de gestión de datos, de archivo de preservación, de gestión de preservación, de acceso y de administración. Igualmente, forma parte del sistema una definición específica de roles y

entidades en la que el Productor son las personas o clientes del sistema que generan o proporcionan la información que debe ser preservada, el Archivo es la entidad que realiza las funciones y servicios necesarios para el control de las operaciones diarias de un sistema OAIS, y, por último, el Consumidor (usuario) que son las personas o clientes que interactúan con los servicios de un sistema OAIS para acceder a la información preservada.

DSpace está diseñado para cumplir con el modelo OAIS. Proporciona las funciones SIP, AIP y DIP. Usa METS como contenedor AIP (Archival Information Package).

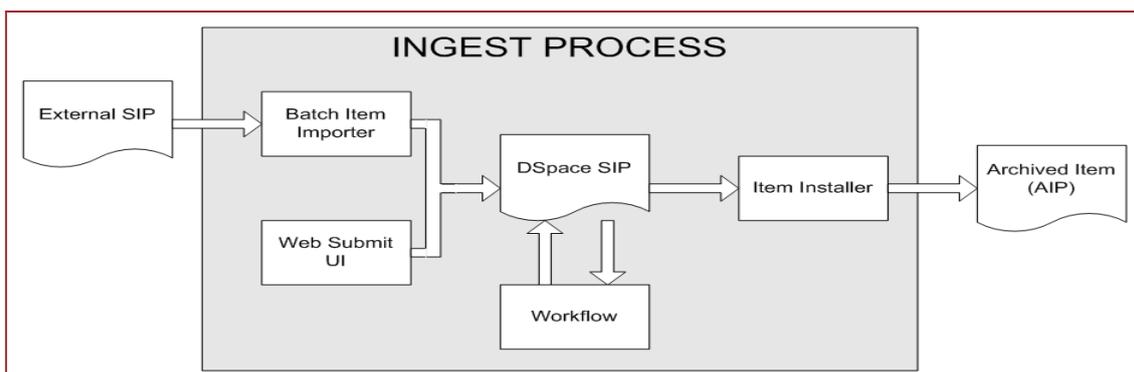


Figura 4. Modelo OAIS en DSpace (Tansley et al., 2005)

METADATOS DE PRESERVACIÓN

PREMIS

El modelo de referencia OAIS fue el adoptado por el estudio realizado en 2002 por el grupo de trabajo *Preservation Metadata Framework*, organizado conjuntamente por la OCLC y la RLG para definir los tipos de información que se deberían asociar a un objeto digital a efectos de preservación y que es el origen del *PREMIS Data Dictionary for Preservation Metadata*. La versión 1.0 del *PREMIS Data Dictionary* fue publicada en 2005, y la actual 2.0 es de marzo de 2008. Este diccionario de datos es una traslación del modelo OAIS a unidades semánticas implementables, bajo la forma de un esquema de metadatos específicos para preservación, y sustentado en encuestas sobre sistemas reales de repositorios de preservación. Al igual que el modelo OAIS, el esquema de metadatos PREMIS no define ninguna estrategia o método de implementación, ni prescribe una tipología de acciones, ni el modo en el que deben recogerse, almacenarse o gestionarse los datos referidos a estas acciones. Es muy conveniente utilizar el esquema PREMIS, dada la riqueza y precisión creciente de las distintas versiones de su diccionario de datos. La no existencia de un repositorio de preservación normalizado no exime de la necesidad de utilizar metadatos de preservación y en concreto PREMIS.

El Diccionario de datos PREMIS define los metadatos de preservación como “la información que utiliza un repositorio para soportar el proceso de preservación digital”. Son ejemplos de actividades de preservación y de cómo pueden soportarlas los metadatos:

- Un recurso debe almacenarse de manera segura para que nadie pueda modificarlo inadvertidamente (o malintencionadamente). La información

de la suma de verificación almacenada como metadatos puede utilizarse para expresar si el archivo almacenado ha variado entre dos momentos en el tiempo.

- Los ficheros deben almacenarse en soportes que puedan leer los ordenadores actuales. Si el soporte está dañado u obsoleto (como los discos blandos de 8" utilizados en los años setenta) puede resultar difícil o imposible recuperar los datos. Los metadatos pueden apoyar la gestión de los distintos soportes mediante el registro del tipo y la edad del soporte de almacenamiento y las fechas en las que se refrescaron los ficheros por última vez.
- Transcurrido un período largo de tiempo, incluso los formatos de fichero más comunes pueden convertirse en obsoletos, lo que significa que las aplicaciones actuales no pueden reproducirlos. Los gestores de la preservación deben emplear *estrategias de preservación* que garanticen que los recursos migren de antiguos formatos a otros nuevos equivalentes (*migración*), o la imitación del antiguo entorno de reproducción en el nuevo hardware y software (*emulación*). Tanto las estrategias de emulación como las de migración requieren metadatos sobre los formatos de los ficheros originales y los entornos de hardware y software que los soportan.
- Las acciones de preservación pueden implicar modificaciones de los recursos originales o cambios en su modo de reproducción. Esto puede poner en duda la autenticidad del recurso. Los metadatos pueden ayudar a soportar la autenticidad del recurso mediante la documentación de la *procedencia digital* de dicho recurso, su cadena de custodia y el historial de cambios autorizados.

El modelo de datos PREMIS define cinco entidades: Entidades Intelectuales, Objetos, Agentes, Acontecimientos y Derechos (figura 5).

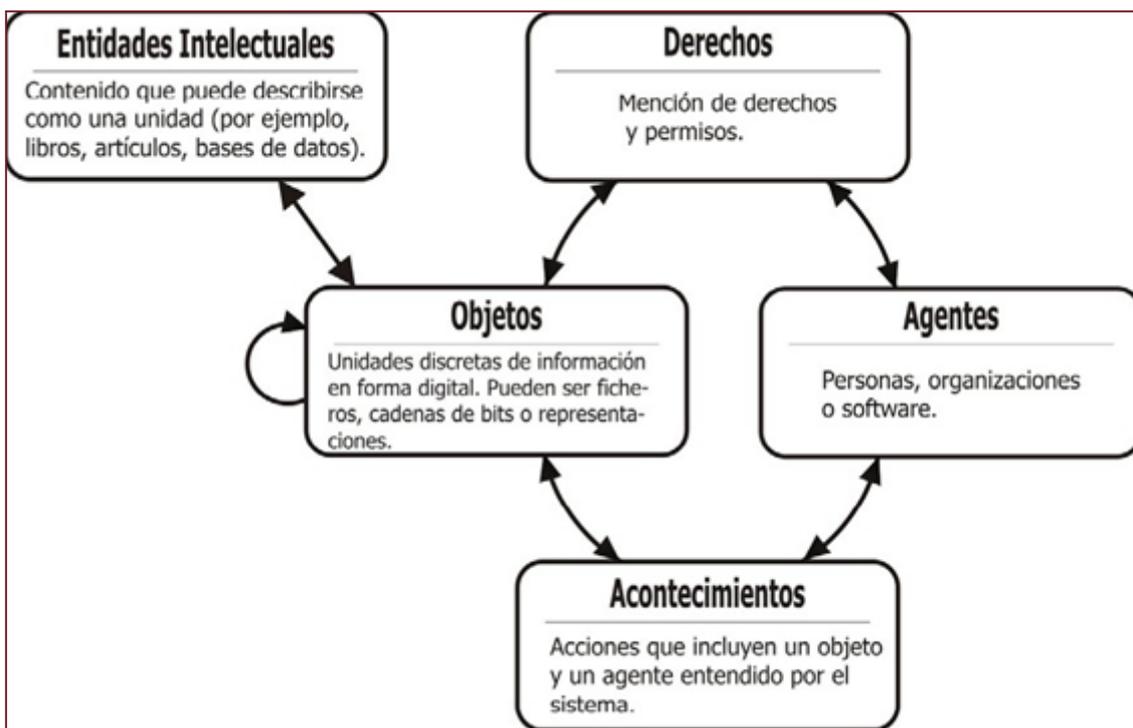


Figura 5. El modelo de datos PREMIS

Se define la *Entidad Intelectual* como “un conjunto de contenido que se considera como una sola unidad intelectual para los propósitos de gestión y descripción: por ejemplo, un libro, un mapa, una fotografía o una base de datos”. PREMIS no define los metadatos correspondientes a las Entidades Intelectuales porque existen muchos estándares de metadatos descriptivos entre los que se puede elegir. Establece que en un sistema de preservación, un objeto debe estar asociado a la entidad intelectual que representa mediante la inclusión de un identificador de dicha entidad en los metadatos del objeto (por ejemplo un ISBN como enlace a la Entidad Intelectual).

Los *Objetos* son unidades de información digital. Son lo que realmente se almacena y gestiona en un repositorio de preservación. La mayor parte de PREMIS se dedica a describir objetos digitales. La información que se puede registrar incluye:

- el identificador único del objeto (tipo y valor),
- fijeza de la información, como la suma de verificación (mensaje cifrado) y el algoritmo utilizado para obtenerla,
- el tamaño del objeto, el formato del objeto, que puede especificarse directamente o mediante un enlace a un registro de formatos, el nombre original del objeto,
- información sobre su creación,
- información sobre los inhibidores,
- información sobre sus propiedades significativas,
- información sobre su entorno (véase más abajo),
- dónde y en qué soporte se almacena,
- información sobre la firma digital,
- relación con otros objetos y otros tipos de entidades.

PREMIS define tres tipos diferentes de objetos y requiere que los implementadores establezcan una distinción entre ellos: objetos *fichero*, objetos *representación* y objetos *cadena de bits*. El objeto fichero es un fichero de ordenador, por ejemplo un fichero PDF o un JPEG. El objeto representación es el conjunto de todos los objetos fichero que se necesitan para reproducir una Entidad Intelectual. Los objetos cadena de bits son subconjuntos de ficheros. El hecho de tener distintos tipos de objetos obliga a pensar en lo que se describe y a ser lo más preciso posible, lo que es importante para el proceso automático.

La *entidad Acontecimiento* agrega información sobre acciones que afectan a los objetos del repositorio. Un registro preciso y fiable de los acontecimientos es crítico para el mantenimiento de la procedencia digital de un objeto lo que, a su vez, es importante para demostrar la autenticidad del objeto. La información sobre los acontecimientos que se puede registrar incluye:

- el identificador único del acontecimiento (tipo y valor),
- el tipo de acontecimiento (creación, ingesta, migración, etc.),
- la fecha y hora en la que ocurrió el acontecimiento,
- la descripción detallada del acontecimiento,
- el resultado codificado del acontecimiento,

- una descripción más detallada del resultado,
- los agentes implicados en el acontecimiento y sus funciones,
- los objetos implicados en el acontecimiento y sus funciones.

Cada sistema de repositorio puede tomar sus propias decisiones sobre los acontecimientos que se registran como parte permanente de la historia de un objeto. PREMIS recomienda que se registren siempre las acciones que modifican un objeto y la entrada para el tipo de acontecimiento (eventType) del Diccionario de Datos proporciona una “lista de entrada” de tipos de acontecimientos importantes con el objetivo de promover que los repositorios registren estos acontecimientos de manera consistente.

Los *Agentes* son actores con funciones en los acontecimientos y en las menciones de derechos. Los agentes pueden ser personas, organizaciones o aplicaciones de software. PREMIS solo define el número mínimo de unidades semánticas necesarias para identificar los agentes puesto que existen varios estándares externos que se pueden utilizar para registrar información más detallada. Un repositorio puede elegir entre utilizar un estándar independiente para registrar información adicional sobre los agentes o utilizar el identificador del agente para apuntar a la información registrada externamente.

El Diccionario de datos incluye:

- un identificador único para el agente (tipo y valor),
- el nombre del agente,
- la designación del tipo de agente (persona, organización, software).

Siempre que se haga referencia a un agente en relación con un acontecimiento o con una mención de derechos, debe registrarse también la función del agente. Cualquier agente puede tener varias funciones. En el modelo PREMIS un repositorio debe asignar un identificador único al agente y debe consignar el identificador del registro de cualquier acontecimiento o mención de derechos en el que se es agente, junto con su función en ese contexto particular. La función de un agente en relación con un acontecimiento o una mención de derechos se consideran una propiedad de la entidad acontecimiento o de la entidad derechos y no del propio agente.

La *entidad Derechos* agrega información sobre los derechos y permisos que son relevantes para preservar los objetos del repositorio. Cada una de las menciones de derechos de PREMIS constata dos cosas: las acciones a las que tiene derecho el repositorio y las bases para la reclamación de ese derecho. La información que puede registrarse en una mención de derechos incluye:

- el identificador único de la mención de derechos (tipo y valor);
- si la base para la reclamación de los derechos es el copyright, una licencia o una ley; información más detallada sobre el estado del copyright; los términos de la licencia o la ley;
- si es aplicable, la(s) acción (es) que permite la mención de derechos; cualquier restricción sobre la(s) acción(es); los derechos otorgados o el período de tiempo durante el que se aplica la mención; el (los) objeto(s) a los que se aplica la mención, los agentes implicados en la mención de derechos y sus funciones.

La mayor parte de la información está diseñada para ser *accionable* (es decir, registrada de una forma controlada que pueda ser ejecutada por un programa de ordenador). La mención de derechos de PREMIS es una consignación de derechos, no un registro de información a partir del que puedan determinarse los derechos. El objetivo de la entidad Derechos de PREMIS es proporcionar información efectiva a los sistemas de los repositorios de preservación.

METS

El formato METS es un formato estándar para codificación y transmisión de metadatos. Está pensado principalmente para el envío de los ficheros, imágenes y objetos multimedia de una biblioteca digital. METS utiliza la estructura de etiquetas tipo XML. Un documento METS está compuesto por 7 secciones principales (no son obligatorias todas ellas):

metsHdr (Mets Header). Información del documento METS: fecha y hora, nombre de la persona que lo crea, etc.

dmdSec (Descriptive Metadata). Descripción del objeto al que se hace referencia en el documento METS. Los metadatos pueden estar definidos de dos maneras distintas: 1. En forma XML en un estándar distinto a METS, podría ser en Dublin Core o en MODS. La descripción en Gredos está hecha en Dublin Core Cualificado (DCQ) pero el formato que se visualiza es DC simple. 2. En cualquier forma textual, binaria pero codificada en Base64.

admSec (Administrative Metadata). Contiene los metadatos administrativos sobre los ficheros que forman el objeto digital y el material fuente del que se han obtenido dichos ficheros. Existen cuatro formas principales de metadatos administrativos disponibles para su utilización en un documento METS:

1. Metadatos técnicos: información relativa a la creación, formato y características de utilización.
2. Metadatos sobre derechos y propiedad intelectual: copyright y licencias.
3. Metadatos sobre la fuente: metadatos descriptivos y administrativos de la fuente de la que se ha derivado el objeto digital.
4. Metadatos sobre el origen digital: información sobre las relaciones origen/destino entre los ficheros del objeto digital.

fileSec (File groups) son los ficheros que comprenden la versión electrónica del objeto digital. Estos ficheros se agrupan en <fileGrp>: por ejemplo podría haber distintos elementos para los archivos master, versiones pdf, versiones reducidas, etc.

structMap (Structural Map) define la estructura jerárquica del objeto y nos permite navegar por él.

smLink (Structural Links) se utiliza para indicar hiperlinks entre el estructural map.

Behaviour Section se utiliza para asociar comportamientos ejecutables con los contenidos del objeto METS.

```

<?xml version="1.0" encoding="utf-8" ?>
<mets:mets xmlns:mets="http://www.loc.gov/METS/"
xmlns:xlink="http://www.w3.org/TR/xlink"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd
http://purl.org/dc/elements/1.1/
http://www.dublincore.org/schemas/xmls/simpledc20020312.xsd">
<mets:dmdSec>
<mets:mdWrap MDTYPE="DC">
<mets:xmlData>
<dc:title>Simulación interactiva de un robot submarino</dc:title>
<dc:creator>Herrero Herrero, Jesús</dc:creator>
<dc:creator>Martín Corral, Sergio David</dc:creator>
<dc:subject>Visualización tridimensional (Informática)</dc:subject>
<dc:subject>Robots</dc:subject>
<dc:subject>Three-dimensional imaging</dc:subject>
<dc:description>Proyecto fin de carrera de Ingeniería Técnica en Informática
de Gestión. El objetivo principal de este trabajo es simular un entorno
submarino 3D, un robot submarino y las interacciones entre ambas
representaciones. El uso de un sistema de estas características permite:
interacción, inmersión, usabilidad, situaciones de emergencia real,
desaparición de distancias, etc. Las ventajas que presentan estas aplicaciones
son la toma de experiencia, la reducción de costes, tiempo y peligros que
podría suponer la inmersión real de un robot submarino.</dc:description>
<dc:description>1. Aplicación; 2. Documentación</dc:description>
<dc:date>2009-07-23</dc:date>
<dc:date>2009</dc:date>
<dc:type>Proyecto fin de carrera</dc:type>
<dc:type>Bachelor Thesis</dc:type>
<dc:format>application/pdf</dc:format>
<dc:format>119 MB</dc:format>
<dc:identifier>http://hdl.handle.net/10366/d10527114</dc:identifier>
<dc:language>Español</dc:language>
<dc:language>spa</dc:language>
<dc:relation>Winzip</dc:relation>
<dc:rights>Acceso abierto / Open Acces</dc:rights>
</mets:xmlData>
</mets:mdWrap>
</mets:dmdSec>
<mets:amdSec ADMID="ADM1">
<mets:rightsMD>
<mets:mdWrap MDTYPE="DC">
<mets:xmlData>
<dc:rights>Acceso abierto /Open Access Licencia Creative Commons
Reconocimiento-No comercial-Sin obras derivadas 3.0 España</dc:rights>
</mets:xmlData>
</mets:mdWrap>
</mets:rightsMD>
</mets:amdSec>
<mets:fileSec>
<mets:fileGrp>
<mets:file ID="FILE1" SIZE="125584042"
CHECKSUM="ccb8f947874c2a8082a421d69731958c" CHECKSUMTYPE="MD5"
ADMID="ADM1">
<mets:fileGrp>
<mets:file ID="FILE1" MIMETYPE="application/pdf" SIZE="75425549"
CHECKSUM="79b004c54d78c23cda6f3b94cd01e035" CHECKSUMTYPE="MD5">
<mets:FLocat LOCTYPE="URL"

```

```

xlink:href="http://gredos.usal.es:80/getblob?id=5024053451059545" />
</mets:file>
</mets:fileGrp>
</mets:fileSec>
<mets:structMap>
<mets:div>
<mets:fptr FILEID="FILE1" />
</mets:div>
</mets:structMap>
</mets:mets>

```

Figura 6. METS en Gredos

USO DE PREMIS CON METS

En lo que se refiere a DSpace, este ha definido un set llamado “Technical Metadata Element” Set para cumplir con la preservación y con las necesidades de gestión del ciclo de vida de la información. Este set se expresa usando el esquema de metadatos de preservación PREMIS: Objeto. DSpace usa el diccionario de datos PREMIS para representar los elementos de metadatos necesarios, aunque no constituye la implementación completa del modelo de datos PREMIS. También hay un requerimiento estructural que facilita el uso de los atributos del elemento “file” (*checksum*, *checksumtype*, *created* y *mimetype*). Estos atributos reflejan elementos en el set Technical Metadata Element. Esto se aplica a la adición de metadatos técnicos a instancias METS, conformando así el perfil SIP (Submission Information Packages). En el nivel de ítem (parecido pero no correspondiendo totalmente con “representación”) DSpace almacena metadatos Dublin Core de procedencia (provenance), algunos son almacenados automáticamente, y otros deben introducirse manualmente. Los metadatos automáticos incluyen nombre del catalogador (más email), fecha, y tamaños más checksums de los bitstreams. Por ejemplo, quién envía el ítem y cuándo, y el tamaño del archivo más checksum.

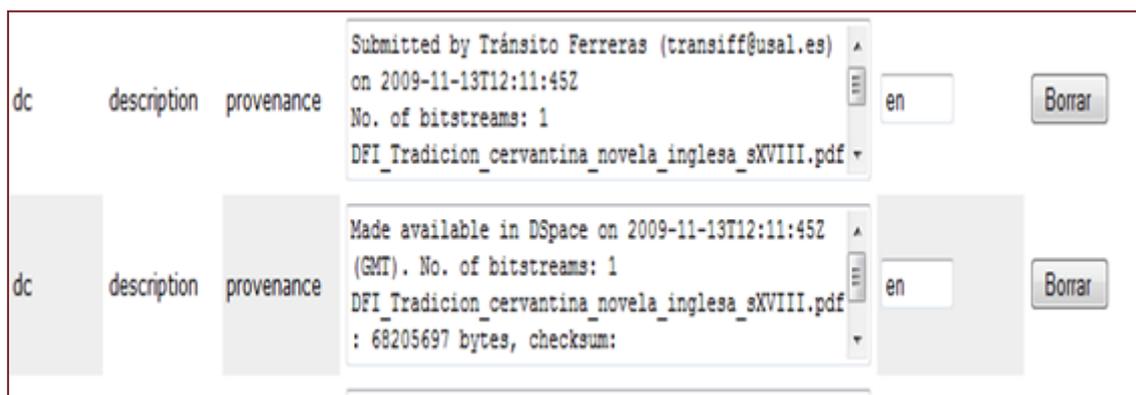


Figura 7. Ejemplo dc.description.provenance en Dspace de Gredos

Además, el sistema almacena la historia de los acontecimientos sobre preservación de acuerdo al modelo Harmony/ABC (en RDF / XML), aunque este sistema es todavía bastante experimental. Además, los archivos log que genera DSpace se podrían construir para contener información de procedencia, de esta forma detallarían quién realiza acciones que afectan al objeto.

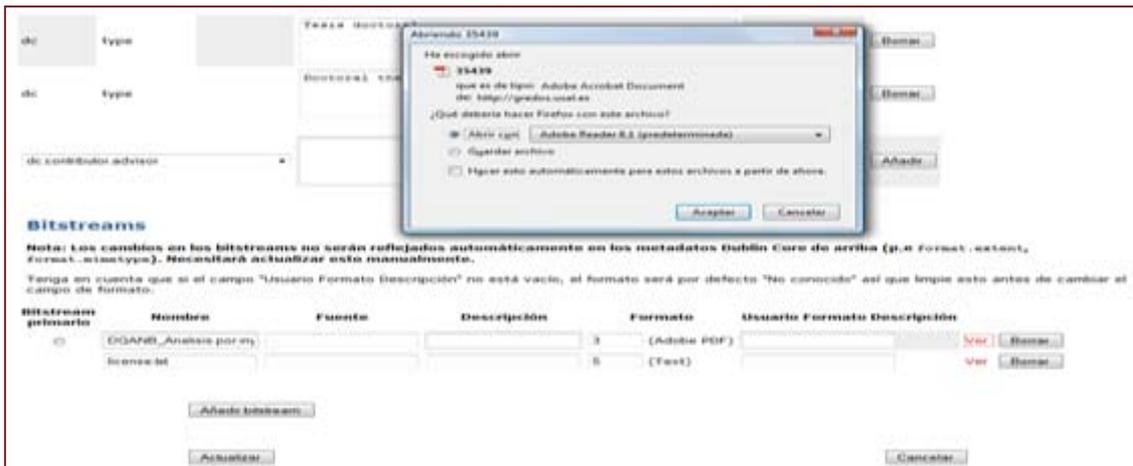


Figura 8. Ejemplo de Gredos. Nivel del bitstream en Dspace.

Los metadatos de un objeto digital pueden realizarse o bien usando PREMIS como esquema básico o bien utilizando PREMIS como una extensión de METS, esta última opción está más extendida, y para lo que pueden utilizarse las recomendaciones *Guidelines for using PREMIS with METS for exchange*. "Using Premis with METS" refleja una serie de perfiles que especifican el uso de los metadatos PREMIS envueltos en un esquema METS y la relación con otras extensiones de metadatos que pueden coincidir total o parcialmente con METS (<http://www.loc.gov/standards/premis/premis-mets.html>). La versión 1.9 de METS hace una recomendación concreta para que se haga así. Cualquier sistema que reciba un paquete de información procedente de un proceso de digitalización, por ejemplo, debe establecer sus requisitos mínimos y obligatorios, o bien haciendo referencia a alguno de los perfiles de METS existentes o bien desarrollando sus propios perfiles. La reciente publicación de *PREMIS in METS Toolbox* (<http://pim.fcla.edu/>) facilita considerablemente la tarea, pues se trata de herramientas de código abierto creadas por el *Florida Center for Library Automation* para la *Library of Congress* en 2009.

Para extraer los metadatos PREMIS con METS en GREDOS, se ha incluido la aplicación XMLUI, además de la JSPUI (que es la que tiene la interfaz) dentro de los webapps del repositorio. Para ver los datos se utiliza una ruta similar a esta:

<http://gredos.usal.es/xmlui/metadata/handle/10366/75282/mets.xml?techMDType=PREMIS>

Sustituyendo los números 10366/75282 por el handle correspondiente se accede a los PREMIS de cada ítem. La aplicación XMLUI es otro tipo de interfaz que además ofrece otras funcionalidades como la generación de Metadatos en diferentes formatos.

El uso de un amdSec con la repetición de subelementos (techMD, etc.) o repitiendo amdSec para cada subelemento METS, depende de las preferencias de cada implementador. Si se hace referencia a amdSec (como es el caso de Gredos) se está referenciando todas las ramas dependientes, con lo cual no es obligatorio repetir amdSec.

Los metadatos técnicos de diferentes esquemas (uno de ellos PREMIS) pueden darse separados en diferentes secciones techMD o en un formato

específico de metadatos que podrían ser incluidos con los metadatos PREMIS bajo `objectCharacteristicsExtension` en `premis: object`.

```
</mets:techMD>
-<mets:techMD ID="techMD_file_141660_PREMIS">
-<mets:mdWrap MDTYPE="PREMIS">
-<mets:xmlData>
-<premis>
-<object>
-<objectIdentifier>
<objectIdentifierType>URL</objectIdentifierType>
-<objectIdentifierValue>
http://gredos2.usal.es/jspui/bitstream/10366%2F75282/1/DGG_SerafinTapia_PrimerasLetras.p
df
</objectIdentifierValue>
</objectIdentifier>
<objectCategory>File</objectCategory>
-<objectCharacteristics>
-<fixity>
<messageDigestAlgorithm>MD5</messageDigestAlgorithm>
<messageDigest>97f1396f1ff64452a01f9d870e749053</messageDigest>
</fixity>
<size>11583394</size>
-<format>
-<formatDesignation>
<formatName>application/pdf</formatName>
</formatDesignation>
</format>
</objectCharacteristics>
<originalName>DGG_SerafinTapia_PrimerasLetras.pdf</originalName>
</object>
</premis>
</mets:xmlData>
</mets:mdWrap>
</mets:techMD>
```

Figura 9. Premis con Mets en Gredos

Así pues, en Gredos se están utilizando 4 elementos:

`<mets:dmdSec>` Contiene los metadatos asociados al ítem en formato Dublin Core.

`<mets:amdSec>` Contiene información acerca de los ficheros asociados al ítem (Por ejemplo el PDF y la licencia).

`<mets:fileSec>` y `<mets:structMap>` Ayudan al usuario a comprender qué archivos están siendo referenciados.

La información completa sobre las directrices a seguir en la implementación de PREMIS con METS son consultables en *Guidelines for using PREMIS with METS for exchange* (<http://www.loc.gov/standards/premis/guidelines-premismets.pdf>).

JHOVE: EXTRACCIÓN AUTOMÁTICA DE METADATOS

La extracción automática de metadatos se puede realizar con la herramienta JHOVE (<http://hul.harvard.edu/jhove/>), desarrollada por la Harvard University permite la identificación automática, validación y caracterización de un conjunto de tipos de objetos digitales.

Gredos no tiene implementada esta herramienta la cual sería muy adecuado utilizar para la extracción automática de metadatos que sirvieran para la preservación. Sería conveniente tener en cuenta, para una posible instalación, que DSpace ofrece un add-ons con una versión abreviada de JHOVE. Con esta herramienta DSpace proporciona control del formato, de la extensión del archivo/bistream y comprueba la presencia de virus.

ALMACENAMIENTO, SOPORTE Y SEGURIDAD

En materia de seguridad de la información es recomendable aplicar la norma ISO/IEC 27002:2005 *Information technology -- Security techniques -- Code of practice for information security management*, que comprende las normas ISO/IEC 17799:2005 e ISO/IEC 17799:2005/Cor.1:2007.

La seguridad de la información se define en esta norma como la preservación de la confidencialidad de forma que se asegure que la información es accesible sólo para aquellos autorizados a tener acceso, la integridad como una garantía de la exactitud e integridad de la información y de los métodos de su procesamiento, y la disponibilidad para asegurar el acceso a los usuarios autorizados cuando estos requieran acceder a la información. Contiene los siguientes apartados:

1. Evaluación y tratamiento del riesgo.
2. Política de seguridad.
3. Organización de la seguridad de la información.
4. Gestión de los activos.
5. Seguridad ligada al personal.
6. Seguridad física y del entorno.
7. Gestión de las operaciones y de las comunicaciones.
8. Control de accesos.
9. Desarrollo y mantenimiento de sistemas.
10. Gestión de incidencias en seguridad de la información.
11. Gestión de la continuidad del negocio.
12. Cumplimiento.

En materia de seguridad, almacenamiento y soporte informáticos, Gredos está gestionado por los servicios informáticos de la Universidad de Salamanca.

La disponibilidad de equipos físicos, la capacidad de almacenamiento y la estructura de comunicaciones de este repositorio es la siguiente:

1. Escáneres: 1 Scanner Book-Eye GSBC; 1 Fujitsu F15530 C; 1 Fujitsu 5110; 1 Epson A3 GT 15000+; 1 Nikon LS 5000 y su aplicación para diapositivas SF 210; 31 escáneres HP modelos 4500 C y 4470 C, 5530 y 5590.
2. Ordenadores personales: 6 Ordenadores de sobremesa CORE DUO 2.2 GHz 2 GB de RAM, 2 ordenadores portátiles Intel Centrino 1.6 Hz, 1 GB de RAM.
3. Discos duros externos: 4 discos externos de 500 GB, 2 discos externos 1 TB.
4. Red académica de la USAL. Servidor: HP ProLiant BL460c. Servidor compacto. 2 vías. 1 x Quad-Core Xeon E5450 / 3 GHz. RAM 4 GB hot-

swap 2.5". HP Dual Port. Disco duro 36 GB hot-swap 2.5" SF- SAS 15000 rpm. StorageWorks 1/8 G2 Tape Autoloader Ultrium 1760. Autocargador de cintas 6.4 TB / 12.8 TB ranuras: 8 LTO Ultrium (800 GB / 1.6 TB). El servidor y las copias de seguridad están ubicados en las instalaciones del servicio informático de la Universidad (CPD).

En cuanto a las copias de seguridad, se realizan de dos formas:

1. Copia de seguridad interna de la aplicación: diaria, incremental en cinta interna del servidor (no se pueden hacer completas por el tipo de cinta que se puede emplear en esta unidad de cinta).
2. Copia del volumen externo donde están los datos: backup completo semanal e incremental el resto de los días en soporte disco. Las copias de seguridad se hacen desde la Unidad de Seguridad y Monitorización del CPD de la USAL.

El servidor y las copias de seguridad están ubicados en las instalaciones del servicio informático de la Universidad de Salamanca (CPD).

Lo recomendable para tener un grado óptimo de preservación ante posibles desastres sería disponer de un repositorio de preservación cuyo servidor estuviera ubicado en un lugar distinto. En este sentido Gredos sigue la normativa del plan general de preservación de la Universidad de Salamanca gestionado por los servicios informáticos de la Universidad (CPD).

Como se ha dicho anteriormente, la preservación digital a largo plazo de Gredos no se va a lograr únicamente a través de las copias de seguridad, la redundancia del sistema de acceso o la replicación byte. Estas acciones puedan ofrecer cierta protección a corto y medio plazo y a veces pueden ser elementos de preservación a largo plazo, pero no son elementos suficientes para la preservación digital. Este punto se debería tener en cuenta en Gredos, para desarrollar un plan de preservación digital a largo plazo, que fuese completo y eficaz.

FORMATOS

Gredos tiene diseñada su propia normativa en cuanto a la gestión de las imágenes y formatos, basada en las diferentes recomendaciones y guías establecidas por distintos proyectos. Ha sido elaborada en función de la calidad de las reproducciones, sus diferentes usos y tipos de acceso, el sistema en el que están integradas para su gestión, incluyendo los metadatos necesarios para ello, y los requisitos de preservación digital. Las recomendaciones seguidas sirven no sólo para proyectos de digitalización, sino que también son aplicables a las características de los objetos que nacen digitalmente, ya que estos se van a ver afectados por los mismos problemas de preservación que los digitalizados.

Cualquier repositorio institucional debe establecer los requisitos técnicos de digitalización en cuanto a la resolución, esquema de bits utilizado para representar un punto (blanco y negro, escala de grises o color), la compresión de los ficheros obtenidos en el proceso de escaneado, los formatos de ficheros, y debe adecuar estos requisitos a las características de los originales. Todo ello redundará en la calidad y fiabilidad del repositorio digital de cara a sus

usuarios, así como en su sostenibilidad. Una fuente de información para el desarrollo de los requisitos de un proyecto de digitalización es *Llevando la teoría a la práctica: Tutorial de digitalización* de imágenes de la biblioteca de la *Cornell University*. En este tutorial se puede encontrar una explicación detallada de todos los conceptos sobre esta área, así como su influencia en la construcción de una colección digital de calidad, y una recopilación de fuentes de información complementarias sobre los requisitos de algunos de los grandes proyectos de digitalización. Otra obra de referencia básica es la 3ª edición (2007) de la *NISO Recommended Practice, A Framework of Guidance for Building Good Digital Collections* (ya citada anteriormente en este trabajo) destinada a servir de guía a instituciones de memoria para la planificación de proyectos de creación de colecciones digitales y a entidades que financian, subvencionan o promueven el desarrollo de colecciones digitales de calidad.

Hay que señalar también que a menudo existe una gran disparidad en los formatos de los ficheros (TIFF, JPEG, JPEG 2000, PDF, etc.) de objetos digitales que se utilizan en los diferentes proyectos, con diferentes fines, esto dificulta la selección de los más convenientes o la detección de efectos colaterales a los formatos elegidos lo cual va a repercutir en la continuidad de los proyectos de digitalización. La *Library of Congress* mantiene un sitio Web de referencia en el que analiza una gran cantidad de formatos (imágenes, texto, audio, video, etc.), sus ventajas, desventajas y utilización más recomendada. Estas recomendaciones van cambiando con la aparición de nuevos formatos y normas incidiendo en la selección del formato más adecuado. En España debe seguirse el trabajo del *Grupo de Trabajo de Colecciones Digitales*, que ha traducido las *Directrices para proyectos de digitalización: de colecciones y fondos de dominio público en particular para aquellos custodiados en bibliotecas y archivos*, resultado del trabajo de un grupo de expertos de IFLA e ICA (*International Council on Archives*) invitados por la UNESCO.

Aunque en Gredos se siguieron las recomendaciones que se han expuesto anteriormente para decidir qué formatos utilizar en el repositorio institucional y qué directrices de digitalización de imágenes adoptar, sería conveniente seguir actualizándose en esta materia con las recomendaciones que vayan apareciendo. Los formatos (tabla 1) y los parámetros de digitalización (tablas 2 y 3) utilizados actualmente en Gredos son los que se indican a continuación.

Texto
<ul style="list-style-type: none"> • PDF
Video
<ul style="list-style-type: none"> • Flash • QuickTime
Imágenes
<ul style="list-style-type: none"> • TIFF (preservación) • JPEG • PDF
Audio
<ul style="list-style-type: none"> • MP3

Tabla 1. Formatos utilizados en Gredos.

Tipo de documento	Definición	Resolución	Tamaño	Formato	Compleción	Nomenclatura
Soporte Papel*	Original color: Color, 24 bits Original b/n: Escala de grises, 8 bits	300-600 ppp	100% / 1:1	TIF	Sin compresión	Seriada
Negativo y Diapositivas B/N	Escala de grises: 8 bits	2400 ppp	100% / 1:1	TIF	Sin compresión	Seriada
Negativo y Diapositivas Color	Color: 24 bits	2400 ppp	100% / 1:1		Sin compresión	Seriada

Tabla 2. Parámetros de digitalización de las imágenes para preservación.

Tipo de documento	Definición	Resolución	Tamaño	Formato	Compleción	Nomenclatura
Soporte Papel*	Original color: Color, 24 bits Original b/n: Escala de grises, 8 bits	150 ppp	100% / 1:1	JPEG o PDF	JPEG: Media	Seriada
Negativo y Diapositivas B/N	Escala de grises: 8 bits	300 ppp	100% / 1:1	JPEG o PDF	JPEG: Media	Seriada

Tabla 3. Parámetros de digitalización de las imágenes para difusión web.

ACCESIBILIDAD

En cuanto a la accesibilidad de los dispositivos de hardware, software y comunicaciones, la utilización de protocolos abiertos asegura la interoperabilidad del repositorio y, por tanto, el intercambio y transferencia de información. El repositorio digital de la Universidad de Salamanca permite que sus contenidos puedan incorporarse a redes de colecciones similares de todo el mundo y ser buscados por servicios OAI como Oaister o por agentes de Internet como Google, etc. A través del interfaz Web, los usuarios podrán localizar fácilmente la información que buscan sin importar el formato: PDF, imágenes digitales en PDF o JPEG, objetos o programas ejecutables. Para

garantizar la persistencia de las URL de los objetos digitales y de los registros se utiliza el *Sistema Handle* del CNRI (<http://www.handle.net/>) para asignar y resolver identificadores persistentes para cada uno de los ítems digitales. Los Handles fueron elegidos, en lugar de URL's, porque permiten la cita de los ítems depositados en Gredos de forma unívoca y persistente a lo largo del tiempo.

SOLUCIONES FRENTE A LAS AMENAZAS

La experiencia de algunos autores nos facilitarán las recomendaciones sobre las estrategias a seguir y cómo aplicarlas a nuestro caso. Las recomendaciones que siguen están basadas en el artículo *Requirements for digital preservation systems* (Rosenthal et al., 2005).

El objetivo primero de un sistema de preservación digital aplicable a Gredos es que la información en él contenida se mantenga accesible para los usuarios por un periodo largo del tiempo. Esto significa que el sistema de preservación debe responder a estas tres propiedades:

- Como mínimo el sistema no puede tener un único punto de fallo, ha de tolerar el fallo de cada uno de los componentes. En general, tendría que poder soportar el fallo de diversos componentes simultáneamente.
- El soporte, el software y el hardware han de correr en el sistema a la vez, a medida que fallan o se tornan obsoletos hay que sustituirlos. El sistema tiene que soportar diversidad de componentes para evitar vulnerabilidades fruto de un solo ámbito, permitir las sustituciones incrementales y evitar el bloqueo del proveedor único.
- La mayoría de los datos de un archivo se usan muy poco. Un sistema que solo detecte errores a través del uso sería vulnerable a una acumulación de errores latentes. El sistema debe procurar hacer comprobaciones regulares en intervalos frecuentes para mantener la probabilidad de fallo en niveles aceptables.

En el sistema del diseño de preservación hay que tener en cuenta las amenazas a las que hay que hacer frente y que tienen que estar identificadas:

- Fallo del soporte
- Fallo de la maquinaria
- Fallo del programa
- Errores en las comunicaciones
- Fallo de los servicios de red
- Obsolescencia de los soportes y de la maquinaria
- Obsolescencia del software
- Error del operador
- Desastre natural
- Ataque externo
- Ataque interno
- Fallo económico
- Fallo organizativo

La degradación del sistema se puede evaluar respondiendo a preguntas tales como: ¿Qué fracción del contenido del sistema se ha perdido de forma

definitiva?, ¿Qué fracción del conjunto de usuarios sufre qué retrasos en el acceso a la fracción afectada pero recuperable del contenido del sistema?

Las estrategias que se pueden aplicar para protegerse de estas amenazas son:

1. Replicación. Es la estrategia más básica y es necesario que las réplicas sean más de una. Algunos archivos de información digital hacen un número fijo de réplicas en servidores distribuidos geográficamente. Además cada réplica puede ser que tenga copias de seguridad.

2. Migración. En sus dos aspectos de cambio de soporte y de cambio de formato representa una de las estrategias para prevenir la obsolescencia del formato y del software. A veces se hace la migración en el momento de la ingesta y entonces se habla de "normalización", decidiendo si se conservará en paralelo el formato original o únicamente el objeto normalizado.

3. Transparencia. Esta cualidad es esencial para que el usuario pueda confiar en un sistema ya que los resultados son siempre futuros. Es de resaltar la importancia de la utilización de código abierto para asegurar el funcionamiento del sistema a largo plazo. Es remarcable la necesidad de compartir la información sobre los casos de pérdida y las mejores prácticas, tomando como modelo y ejemplo el sistema establecido por la NASA para informar de las incidencias en la navegación aérea.

4. Diversidad. Los sistemas que no admiten la diversidad son mucho más vulnerables que aquellos que lo hacen. Por lo tanto un sistema de preservación debería de conllevar diversidad a diferentes niveles:

- soportes fuera de línea,
- dispersión geográfica, por ejemplo DAITSS (Dark Archive In The Sunshine State), BL (British Library) , LOCKSS (Lots of Copies Keep Stuff Safe),
- diversidad de maquinaria y de proveedores, por ejemplo BL (British Library),
- diversidad de programas y proveedores por ejemplo BL (British Library)
- diversidad de administradores, por ejemplo BL (British Library), LOCKSS,
- diversidad de fuentes de financiación.

5. Auditoría. Es necesario establecer mecanismos de control que verifiquen en diferentes fases del proceso (Ingesta, Archivo, Difusión) y mediante técnicas diversas que la información se mantiene inalterada. Las comprobaciones regulares para evitar la acumulación de errores latentes dan confianza a los que financian el sistema de preservación digital.

6. Contención de los costes. Uno de los peligros más grandes para garantizar la puesta en marcha y el mantenimiento de un sistema de preservación es el coste porque pocas instituciones u organizaciones pueden dedicar financiación a usos futuros y por la dificultad en mostrar los resultados de las inversiones realizadas. Algunas propuestas para afrontar esta cuestión:

- el uso de licencias Creative Commons ayuda a reducir los costes asociados a la obtención de permisos que en muchas ocasiones son prohibitivos. Gredos está utilizando ya este tipo de licencias.
- la automatización de las tareas de ingesta incluida la extracción de metadatos,
- en cuanto a la adquisición y la sustitución regular del programa y de las maquinaria se considera más relevante la cantidad de copias que la calidad de la maquinaria. El sistema de preservación no requerirá una respuesta muy rápida,

- por lo que respecta a los costes operativos se insiste que haya administraciones separadas para evitar que posibles errores afecten a más de una réplica,
- hay que tener en cuenta el coste de las migraciones masivas de formato,
- también hay que prever los costes de mantener y asegurar las restricciones en el acceso,
- un elemento a favor de los sistemas de preservación es la lentitud con que se pueden llevar a cabo las operaciones y que permiten ahorros de hardware y de administración.

El resumen de requerimientos que tendría que tener el sistema de preservación digital para el repositorio de la Universidad de Salamanca sería:

- Disponer de un modelo de amenazas y cómo protegerse de ellas.
- Explicar cómo se crean y administran las réplicas y cómo se detectan y reparan los daños.
- Tener políticas y mecanismos implementados para proteger la propiedad intelectual. Específicamente se debe poder mostrar a quién pertenecen los derechos y bajo qué condiciones se encuentra el contenido.
- Hacer públicas las interfaces, en particular las especificaciones SIP y DIP.
- Hacer públicas las políticas de acceso a su código fuente y cómo se preservará este.
- Hacer público quién hará las auditorias, cómo se harán y a quién se informará del resultado.
- Hacer público el tratamiento de incidencias, a quién se informa y de qué forma.

Moore et al. (2005) recomiendan como una buena solución la utilización de la tecnología Grid, como en las colecciones de NARA (archivos gubernamentales de Estados Unidos), con ella se minimizan los riesgos de la pérdida de datos y se puede utilizar maquinaria corriente. En el caso de NARA la copia principal y el catálogo completo de metadatos están en un lugar diferente que la réplica que se encuentra en la Universidad de Maryland con el fin de mejorar el acceso, balancear la carga y poder recuperar en caso de desastre, también mantienen un archivo oscuro en otro centro (SDSC), sin acceso por parte de los usuarios pero con la copia completa. Con esta arquitectura se toman las medidas oportunas para prevenir riesgos de pérdida de datos contra las siguientes amenazas: fallo del soporte, error sistémico del software o del hardware / proveedor, error operacional, desastre natural, usuario malhechor.

Otro ejemplo es la British Library, en la que se ha implementado una arquitectura alternativa con localizaciones autónomas. Las principales características técnicas de este sistema es que se trata de un sistema de depósito permanente: longevidad de los datos, coste bajo y escalabilidad a través del tiempo, las tecnologías y los proveedores. Las soluciones a las amenazas se encuentran en réplicas en localizaciones autónomas, requerimientos bajos de ingeniería y la posibilidad de escalar a través del tiempo y de diferentes tecnologías.

CONCLUSIONES

Antes de finalizar la redacción de este trabajo ha aparecido una importante noticia en los medios de comunicación en relación al proyecto de digitalización del patrimonio cultural europeo. Este proyecto supondrá, sin lugar a dudas, el desarrollo de técnicas de preservación digital a largo plazo que son irremediablemente necesarias para asegurar la accesibilidad y la perdurabilidad del material digitalizado a lo largo del tiempo. El *Comité de Sabios de la Unión Europea* para la digitalización del patrimonio cultural europeo propondrá, antes de que acabe el año 2010, un informe con las líneas a seguir en materia de digitalización de la cultura, de protección de la creación y de asociación público-privada para la digitalización del material cultural. El objetivo de este Comité es presentar recomendaciones que permitan a la Unión Europea dar una respuesta coordinada a los desafíos que representa la digitalización de la cultura y su aportación a la sociedad del conocimiento en los siguientes ámbitos: los costes y la financiación de la digitalización, la identificación de los mejores modelos de acceso a la cultura, y los mecanismos e instrumentos que garanticen la preservación de los recursos a largo plazo. Es este último punto el que nos interesa como conclusión y ejemplo de la relevancia del tema presentado en este trabajo.

Este proyecto europeo nos enfrenta a una clara necesidad: la de asegurar que la información digital accesible hoy al público lo siga siendo a lo largo del tiempo. Este acceso permanente y la conservación de los recursos digitales se alcanzarán investigando, desarrollando y poniendo en práctica técnicas en preservación digital a largo plazo.

Recordemos aquí, citando esta vez a Bollacker (2010), que la preservación digital es el conjunto de estrategias, procesos y técnicas que dan respuesta a los problemas que plantea la conservación de los materiales digitales y de los medios (hardware y software) que se emplean para su almacenamiento y consulta, y que están derivados fundamentalmente de la obsolescencia provocada por la rápida renovación tecnológica y por la inestabilidad de los soportes. Estas técnicas son muy variadas y responden a diferentes situaciones y líneas estratégicas (copias de seguridad, copia de datos en soportes durables, migración, replicación, emulación, transferencia de datos, etc.), aunque, en general, están destinadas a mantener los objetos digitales y sus características de acceso a largo plazo.

Desde el punto de vista del usuario es necesario hacer hincapié en que la preservación digital, además de su objetivo primario, el de poder utilizar a largo plazo la información digital o digitalizada, presenta una característica que debe estar siempre presente como es la posibilidad de consultar la información preservada, hasta el punto de que la consulta es uno de los elementos básicos del modelo de preservación digital.

En este sentido, uno de los mayores problemas que se han planteado para el proyecto de *Google Books* es que en él no está previsto, de modo explícito, la preservación a largo plazo de los contenidos. Este es un aspecto que no debe dejarse de lado y quizá la mejor manera de expresar el objetivo que se persigue sea afirmando que preservación y acceso son las dos caras de la

misma moneda y para que ello sea una realidad hay que intentar que los procesos que conducen a la preservación también permitan el acceso, lo cual debe realizarse desde el momento de la creación del objeto digital y mediante el uso de diferentes esquemas de metadatos. Un ejemplo de ello puede encontrarse en CLOCKSS (*Controlled LOCKSS*), organización sin fines de lucro formada por editores científicos y bibliotecas especializadas con el objetivo de construir un archivo de preservación a largo plazo para publicaciones científicas en la Web.

Lo que será necesario es, en la medida de lo posible, procurar que esos procesos se retroalimenten unos a los otros, de modo que el esfuerzo en la creación de objetos digitales no redunde en una dejación de las funciones específicas de la preservación, todo lo cual es perfectamente viable si dentro del flujo de trabajo de creación de nuevos objetos digitales se prevé la obtención de los mismos.

Desde el punto de vista de la organización y financiación de las políticas de preservación, deberán ser las instituciones, organismos y comunidades interesadas las que se comprometan en el desarrollo, participación y puesta en práctica de proyectos sobre preservación digital que aseguren la perdurabilidad y accesibilidad de la cultura y de la información digital para las generaciones venideras.

El repositorio institucional de la Universidad de Salamanca, es un servicio a la comunidad universitaria, extensible a toda la ciudadanía, que ofrece el acceso libre y gratuito a la producción digital de la Universidad, con fondos de elevado interés patrimonial y científico. Por ello debería ser una prioridad institucional que el repositorio, reflejo del trabajo de sus investigadores y reflejo del rico patrimonio cultural de esta Universidad, pudiera seguir ofreciendo, pasadas unas décadas, el acceso y la consulta a *Gredos*.

El plan de preservación digital para *Gredos* tiene su base primera, su cimentación, en los aspectos técnicos que se están ya desarrollando y poniendo en práctica en la gestión diaria del repositorio, aunque se necesita hacer más en este sentido. Es necesario el compromiso y el apoyo de la institución para desarrollar la política de preservación en su totalidad. Como hemos visto no podemos volver la espalda a un problema latente como es el de asegurar el acceso a la documentación digital a largo plazo, y por lo tanto asegurar el acceso a la información, a la cultura y al patrimonio de una institución.

BIBLIOGRAFÍA

Agenjo Bullón, X. y, Hernández Carrascal, F. (2010). La biblioteca virtual: función y planteamiento. [Preprint]. Disponible en: <http://eprints.rclis.org/18056/> [Consultado el 18 de junio de 2010].

Bates, M. et al. (2006). Digital lifecycles and file types: final report. JISC; Loughborough University. Disponible en: <http://hdl.handle.net/2134/1793> [Consultado el 15 de junio de 2010].

Beagrie, N. (2004). Digital Information Will Never Survive by Accident. SAP INFO. Disponible en: http://sap.info/archive/interviews/interviews_digital_information_will_never_survive_by_accident_02.08.2004.html. [Consultado el 15 de junio de 2010].

Bevan, S.J. (2007). Developing an institutional repository: Cranfield QUEprints – a case study. *OCLC Systems & Services*, 23 (2), 170-182.

Biblioteca de la Universidad de Cornell (2005). Llevando la teoría a la práctica: Tutorial de digitalización. Disponible en: <http://www.library.cornell.edu/preservation/tutorial-spanish/contents.html>. [Consultado el 20 de mayo de 2010].

Bollacker, Kurt D. (2010) Avoiding a Digital Dark Age : Data longevity depends on both the storage medium and the ability to decipher the information. *American Scientist* 98 (2), 106. Disponible en: <http://www.americanscientist.org/issues/pub/2010/3/avoiding-a-digital-dark-age> [Consultado el 20 de mayo de 2010]

Caplan, P. (2009). Entender PREMIS. Madrid, Subdirección General de Coordinación Bibliotecaria. Disponible en: http://www.loc.gov/standards/premis/UnderstandingPREMIS_espanol.pdf. [Consultado el 20 de mayo de 2010].

Chavez, R., et al. (2007). Services make the repository. *Journal of Digital Information*, 8(2). Disponible en: <http://journals.tdl.org/jodi/article/viewArticle/195/179> [Consultado el 21 de junio de 2010].

CLOCKSS. Disponible en: <http://www.clockss.org/clockss/Home> [Consultado el 15 de junio de 2010].

Consultative Committee for Space Data Systems (2002). Reference Model for an Open Archival Information System (OAIS): CCSDS 650.0-B-1: Blue Book: Recommendation for Space Data System Standards. Disponible en: <http://public.ccsds.org/publications/archive/650x0b1.pdf>. [Consultado el 15 de junio de 2010].

Digital Preservation Coalition (2008). Digital Preservation Handbook. Disponible en <http://www.dpconline.org/vendor-reports/download-document/299-digital-preservation-handbook.html> [Consultado el 31 de mayo de 2010].

DRIVER Guidelines, V.2.0. (2008). Disponible en: http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf . [Consultado el 31 de mayo de 2010]

Dublin Core Metadata Initiative. (2007). Disponible en: <http://dublincore.org/> . [Consultado el 20 de mayo de 2010].

Dublin Core Metadata Initiative. (2008, January 1). *Metadata terms*. Disponible en: <http://dublincore.org/documents/2008/01/14/dcmi-terms/> . [Consultado el 20 de mayo de 2010]

España. Ministerio de Cultura (2005). Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos. Madrid: Dirección General del Libro, Archivos y Bibliotecas.

European Commission. Directorate-General for the Information Society (2002). Digicult report: technological landscapes for tomorrow's cultural economy; unlocking the value of cultural heritage. Luxembourg: Office for Official Publications of the European Communities. Disponible en: <http://digicult.salzburgresearch.at/> [Consultado el 15 de junio de 2010].

A Framework of Guidance for Building Good Digital Collections : A NISO recommended practice, 3rd Edition. (2007). Disponible en: <http://www.niso.org/publications/rp/framework3.pdf>. [Consultado el 15 de junio de 2010].

Gallart, N., y Keefer, A. Preservació de dipòsits digitals (2007). Disponible en: <http://www.recercat.net/bitstream/2072/4078/1/0703Preservacio.pdf>. [Consultado el 31 de mayo de 2010].

García Peñalvo, F.J., Merlo Vega, J.A., Ferreras Fernández, T., et al. (2010). Qualified Dublin Core Metadata Best Practices for GREDOS. *Journal of Library Metadata*, 10 (1), 13-36.

Guidelines for using PREMIS with METS for exchange. (2008). Disponible en: <http://www.loc.gov/standards/premis/guidelines-premismets.pdf> [Consultado el 15 de junio de 2010].

Hitchcock, S., Brody, T., Hey, J. and Carr, L. (2007) Digital Preservation Service Provider Models for Institutional Repositories: Towards Distributed Services. *DLib Magazine*, 13 (5/6). Disponible en: <http://www.dlib.org/dlib/may07/hitchcock/05hitchcock.html> [Consultado el 21 de junio de 2010].

History of the Dublin Core Metadata Initiative. Disponible en: <http://dublincore.org/about/history/> [Consultado el 20 de mayo de 2010].

i2010: bibliotecas digitales [COM (2005) 465 final – Diario Oficial C 49 de 28.2.2008]. Disponible en: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52005DC0465:ES:NOT> [Consultado el 1 de mayo de 2010].

ISO/IEC 27002:2005 (2007). Information Technology – Security techniques – Code of practice for information security management.

JHOVE - JSTOR/Harvard Object Validation Environment website Disponible en: <http://hul.harvard.edu/jhove/>. [Consultado el 19 de junio de 2010].

Lynch, C. (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL Bimonthly Report* 226. Disponible en: <http://www.arl.org/newsltr/226/ir.html>. [Consultado el 20 de mayo de 2010].

Lynch, C., y Lippincott, J.K. (2005). Institutional Repositories deployment in the United States as of early 2005. *D-Lib Magazine*, 11 (9). Disponible en: <http://www.dlib.org/dlib/september05/lynch/09lynch.html> [Consultado el 20 de mayo de 2010].

Marcia Zeng, Jaesun Lee and Allene Hayes (2007). Survey on metadata decisions. Disponible en: <http://www.ifla.org/en/news/survey-on-metadata-decisions-for-digital-libraries>. [Consultado el 15 de junio de 2010].

METS. Disponible en: <http://xml.coverpages.org/mets.html>. [Consultado el 20 de mayo de 2010].

METS: Introducción y tutorial (2008). Disponible en: http://www.loc.gov/standards/mets/METSOverview_spa.html. [Consultado el 20 de mayo de 2010]

OCLC/RLG. Trusted Digital Repositories (2002). Disponible en www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf [Consultado el 31 de mayo de 2010].

PREMIS in METS Toolbox (2009). Disponible en <http://pim.fcla.edu/> [Consultado el 15 de junio de 2010].

PREMIS Working Group (2008). Data Dictionary for Preservation Metadata, version 2.0. Dublin (Ohio): OCLC; Mountain View (California): RLG. <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>

PREMIS Working Group (2004). Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in The Cultural Heritage Community. Disponible en: <http://www.oclc.org/research/projects/pmwg/surveyreport.pdf> . [Consultado el 1 de mayo de 2010].

Rodríguez-Gairín, J.M. DSpace, y Sulé Duesa, A. (2008). DSpace: un manual específico para gestores de la información y la documentación. *BIB, textos universitarios de biblioteconomía i documentació*, 20. Disponible en: <http://www.ub.es/bid/20rodri2.htm> [Consultado el 15 de junio de 2010].

Rosenthal, D.S.H., Robertson, T., Lipkis, T., Reich, V. y, Morabito, S. (2005). Requirements for digital preservation systems. *D-Lib magazine*, 11 (11). Disponible en: <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html> [Consultado el 12 de junio de 2010]

Rothenber, J. (1995) ¿Son perdurables los documentos digitales? *Investigación y Ciencia*, marzo.

Rothenberg, J. (1999). Ensuring the Longevity of Digital Information. Santa Monica, California: RAND. Disponible en: <http://www.clir.org/PUBS/archives/ensuring.pdf> [Consultado el 20 de mayo de 2010].

Salo, D., y Donohue, T. (2006). DSpace how-to guide: tips and tricks for managing common DSpace chores. Disponible en: <http://hdl.handle.net/1920/1045>. [Consultado el 15 de junio de 2010]

Sánchez, S., y Melero, R. (2006). La denominación y el contenido de los Repositorios Institucionales en Acceso Abierto: base teórica para la “Ruta Verde”. Disponible en: <http://eprints.rclis.org/6368/> [Consultado el 20 de mayo de 2010].

UNESCO (2003). Carta para la preservación del patrimonio digital. Paris: UNESCO. Disponible en: http://portal.unesco.org/ci/en/files/13367/10676067825Charter_es.pdf/Charter_es.pdf [Consultado el 20 de mayo de 2010].

UNESCO (2003). Directrices para la preservación del patrimonio digital. Paris: UNESCO. Disponible en: <http://unesdoc.unesco.org/images/0013/001300/130071s.pdf> [Consultado el 20 de mayo de 2010].

Waller, M., y Sharpe, R. (2006). Mind the gap: assessing digital preservation needs in the UK. Digital Preservation Coalition. Disponible en: <http://www.dpconline.org/docs/reports/uknamindthegap.pdf> [Consultado el 31 de mayo de 2010].

