

Archivos y metadatos de conservación: estado del arte y propuesta metodológica

Alejandro Delgado Gómez

Servicio de Archivo y Bibliotecas, Ayuntamiento de Cartagena (España)

Resumen

Se analizan los metadatos archivísticos de conservación, a los que se ha dedicado menos atención que a los metadatos descriptivos y que, sin embargo, son también necesarios para gestionar datos en un entorno digital. En primer lugar, el presente texto proporciona definiciones bien aceptadas de metadatos de conservación. En segundo lugar, revisa algunos proyectos existentes y, a partir de ellos, sugiere una propuesta de implantación en el entorno archivístico. Para ello, se examinan un modelo de referencia ampliamente aceptado —OAIS (Reference Model for an Open Archival Information System)— y cinco experiencias: el proyecto Preservation Metadata for Digital Collections de la National Library of Australia (NLA), la red Networked European Deposit Library (NEDLIB), el proyecto CURL Exemplars in Digital Archives (CEDARS), el proyecto conjunto del Online Computer Library Center (OCLC) y el Research Libraries Group (RLG) PREMIS (PREservation Metadata: Implementation Strategies), y el proyecto de la New Zealand National Library Metadata Standards Framework-Preservation Metadata. En el trabajo inicial, estos proyectos se describieron y se evaluaron cualitativamente, tomando como criterios los de modularidad, extensibilidad, refinamiento, multilingüismo, interoperabilidad y reusabilidad, así como el de existencia de una gramática formal y de una gramática formalizada. Se tabularon los elementos de estos proyectos para facilitar una comparación entre ellos. El resultado fue una serie de elementos de metadatos ya existentes, así como ejemplos y una discusión acerca de su viabilidad en dos entornos: el de las bases de datos relacionales y el de los lenguajes de marcado. En el primer entorno, se sugería una definición de campos y sus posibilidades de utilización. En el segundo, se propuso una instancia de documento XML (Extensible Markup Language), susceptible de ser utilizada como base para la generación de metadatos de conservación en archivos digitales. Esta instancia sirvió para generar automáticamente varias gramáticas (schemas y DTD). Se tomaron en consideración las posibilidades de combinación de las mejores características de ambos entornos. Por último,

se propusieron, a partir de ejemplos de software disponibles, opciones para que la utilización de metadatos de conservación en archivos digitales resultara viable.

Palabras clave: Archivos digitales. Metadatos de conservación. Modelo de referencia. Open Archival Information System (OAIS).

Abstract

A substantial part of the effort developed in order to define sets of archival metadata focuses on descriptive metadata; however, other types of metadata exist, needed in order to manage data in a digital environment. One of them is preservation metadata, studied throughout different projects. First, this paper suggests well-accepted definitions for preservation metadata. Secondly, it reviews some existing projects, and from them suggests a proposal for implementation in an archival environment. In order to reach this aim, it investigates a broadly accepted reference model —OAIS (Reference Model for an Open Archival Information System)— as well as five experiences: the National Library of Australia (NLA) Preservation Metadata for Digital Collections project, Networked European Deposit Library (NEDLIB), the CURL Exemplars in Digital Archives (CEDARS) project, the Online Computer Library Center (OCLC) and Research Libraries Group (RLG) joint project PREMIS (PREservation Metadata: Implementation Strategies), and the New Zealand National Library Metadata Standards Framework-Preservation Metadata project. The initial effort described and qualitatively evaluated these projects, on the basis of the following criteria: modularity, extensibility, refinement, multilinguism, interoperability and reuse, as well as on the existence of both a formal and a formalized grammar. Data elements were tabulated in order to make comparison easier. The result was a set of pre-existing metadata elements, as well as examples and a discussion about its feasibility in two environments: relational databases and markup languages. For the first environment, a definition of fields and possibilities for its use were suggested. In the second one, an XML document was proposed, capable of being used as the basis for the creation of preservation metadata for digital archives. This document was used to automatically generate several grammars (schema and DTD). Possibilities of merging the features of these environments were taken into consideration. Finally, on the basis of the available software, some options to make feasible the use of preservation metadata were proposed.

Keywords: Digital archives. Open Archival Information System (OAIS). Preservation metadata. Reference model.

1. Introducción

La mayor parte de los esfuerzos desarrollados para la definición de series de metadatos archivísticos se enfoca sobre los metadatos descriptivos y de des-

cubrimiento de recursos, así como sobre su presencia en los principales buscadores; sin embargo, existen otros tipos de metadatos, no tan populares como los descriptivos, pero también necesarios para gestionar datos en un entorno digital. Uno de estos tipos está formado por los metadatos de conservación, que han sido estudiados en los últimos años, sin que la mayor parte de las propuestas hayan tenido éxito generalizado, con la potencial excepción de PREMIS (PREservation Metadata: Implementation Strategies). En primer lugar, el presente texto proporciona definiciones bien aceptadas de metadatos de conservación. En segundo lugar, intenta revisar algunos proyectos existentes y, a partir de ellos, sugerir una propuesta de implantación en el entorno archivístico. Para ello, se examinan un modelo de referencia ampliamente aceptado —OAIS (Reference Model for an Open Archival Information System) (1)— y cinco experiencias: el proyecto Preservation Metadata for Digital Collections de la National Library of Australia (NLA), del año 1999, la red Networked European Deposit Library (NEDLIB), del año 2000, el proyecto CURL Exemplars in Digital Archives (CEDARS), del año 2002, el citado proyecto conjunto del Online Computer Library Center (OCLC) y el Research Libraries Group (RLG) PREMIS (PREservation Metadata: Implementation Strategies), también del año 2002, y el proyecto de la New Zealand National Library Metadata Standards Framework-Preservation Metadata (Revised), del año 2003. Todos ellos se declaran en mayor o menor grado dependientes del modelo de referencia Reference Model for an Open Archival Information System (OAIS); además, el proyecto PREMIS, que ha alcanzado un notable grado de solidez y que agrupó en su comienzo a representantes de algunos de los proyectos anteriores, lleva a cabo un análisis comparativo tanto de OAIS como de esfuerzos precedentes en relación con el modelo de referencia y en las relaciones de unos con otros. Por otra parte, su serie de metadatos recoge aquellos que ya se definieron como necesarios en tales esfuerzos precedentes. Tres de estos proyectos —CEDARS, NEDLIB y PREMIS— se basan en el modelo de referencia OAIS, aunque llegan a conclusiones diferentes; los otros dos proyectos —NLA y NZNL— son independientes, aunque NZNL es explícitamente dependiente de NLA. Además, uno de los proyectos —PREMIS— pretende integrar experiencias anteriores. En el trabajo inicial, estos proyectos se describieron y se evaluaron cualitativamente, tomando como criterios los de modularidad, extensibilidad, refinamiento, multilingüismo, interoperabilidad y reusabilidad, así como el de existencia de una gramática formal y de una gramática formalizada. Se tabularon los elementos de estos proyectos para facilitar una comparación entre ellos. El resultado fue una serie de elementos de metadatos ya existentes, así como ejemplos y una discusión acerca de su viabilidad en dos entornos: el de las bases de datos relacionales y el de los lenguajes de marcado. En el primer entorno, se sugería una definición de campos y sus posibilidades de utilización. En el segundo entorno, se propuso una instancia de documento XML (Extensible

Scire. 11: 1 (en.-jun. 2005) 83-101. ISSN 1135-3761.

Markup Language), susceptible de ser utilizada como base para la generación de metadatos de conservación en archivos digitales. Esta instancia sirvió para generar automáticamente varias gramáticas (schemas y DTD). Se tomaron en consideración las posibilidades de combinación de las mejores características de ambos entornos. Por último, se propusieron, a partir de ejemplos de software disponibles, opciones para que la utilización de metadatos de conservación en archivos digitales resultara viable (2). El presente texto incluye un apartado de conclusiones a partir de los trabajos anteriores.

2. Definición de metadatos de conservación

En primer lugar, definiremos *metadatos* de acuerdo con una de sus acepciones más genéricas y, por tanto, susceptible de aplicación en diferentes entornos. De acuerdo con la primera parte de la norma ISO 11179, metadatos son “datos que describen otros datos o procesos” (ISO/IEC, 2004, p. 14). Esta definición tiene la indiscutible ventaja de que solventa la interminable discusión acerca de qué sean metadatos, al considerarlos simplemente datos con un objeto determinado.

La definición que proporciona el proyecto PADI (Preserving Access to Digital Information) es la siguiente (National Library of Australia, 2003):

esencial para asegurar la accesibilidad a largo plazo es el desarrollo de maneras estructuradas de describir y registrar la información necesaria para gestionar la conservación de recursos digitales. Esto se llama comúnmente metadatos de conservación. En contraste con los esquemas de metadatos descriptivos (por ejemplo, MARC, Dublin Core), que se utilizan en el descubrimiento e identificación de objetos digitales, los metadatos de conservación se consideran a veces como una sub-serie de los metadatos administrativos, asistiendo a la gestión de la información, y de los metadatos técnicos, asistiendo al acceso actual al contenido digital. Los metadatos de conservación están dirigidos a almacenar detalles técnicos sobre el formato, estructura y uso del contenido digital, la historia de todas las acciones llevadas a cabo sobre el recurso, incluyendo cambios y decisiones, la información sobre la autenticidad, como las características técnicas o la historia de la custodia, y la información sobre responsabilidades y derechos aplicables a las acciones de conservación.

El modelo de referencia OAIS define metadatos como “datos acerca de datos”, y la descripción de la conservación como “la información que es necesaria para la adecuada conservación de la información de contenido, y que se puede categorizar en información sobre la procedencia, referencia, autenticidad y contexto” (Consultative Committee for Space Data Systems, 2002).

Por su parte, la National Library of Australia explica su proyecto en los siguientes términos (Phillips et al., 1999):

Es importante darse cuenta de que nuestra propuesta de serie de metadatos de conservación tiene la intención de ser una definición de la información que consideramos

necesaria para gestionar la conservación de colecciones digitales. Esto significa que es un modelo de salida de datos, no un modelo de entrada de datos. Esto es, la información que queremos extraer de un sistema de metadatos, no necesariamente qué datos debieran introducirse, por quién y en qué momento; no es de interés por sí mismo el modo en que los metadatos debieran asociarse con lo que se está describiendo. Creemos que este modelo debiera ser aplicable a muchas implantaciones, que quizá decidan registrar esta información de diversas maneras. Este modelo simplemente dice: “haga lo que haga, esto es lo que usted debe proporcionar para que nosotros podamos gestionar la conservación”. También es importante hacer notar que nos estamos enfocando solamente sobre los requisitos de conservación. La serie de metadatos propuesta no pretende tratar con nada más. Reconocemos que en cualquier sistema de implantación es probable que exista un solapamiento entre metadatos registrados para diferentes propósitos. Al enfocarnos sobre la información que necesitamos extraer del sistema para gestionar la información, dejamos de lado la cuestión de si elementos particulares pueden estar ya incluidos en, digamos, otros metadatos administrativos o de descubrimiento de recursos.

El proyecto CEDARS argumenta del siguiente modo la necesidad de los metadatos de conservación:

En el fondo, los metadatos de conservación son todos los diversos tipos de datos que nos permitirán la recreación e interpretación de la estructura y el contenido de los datos digitales que se hayan conservado (Ludäscher, Marciano & Moore, 2001). Definidos de este modo, está claro que los metadatos de conservación tienen que soportar cierto número de funciones relacionadas, pero distintas. Lynch (1999), por ejemplo, dice que dentro de un depósito digital, ‘los metadatos acompañan y hacen referencia a cada objeto digital y proporcionan información asociada descriptiva, estructural, administrativa, de gestión de derechos y de otros tipos’. La amplia gama de funciones que los metadatos de conservación pretenden satisfacer significa que la definición de normas de metadatos no es una tarea simple, y que muchos de los esquemas publicados actualmente son relativamente complejos. La situación se complica adicionalmente por la percepción de que diferentes estrategias de conservación digital o de tipos de información digital requerirán el soporte de diferentes tipos de metadatos (CEDARS, 2002).

El proyecto NEDLIB, tanto en su glosario como en su lista de términos, recurre a definiciones bien aceptadas de metadatos y de conservación. Así, metadatos son (Borbinha et al., 2000, pp. 28-29):

1. Información estructurada, contenida quizá en una cabecera adjunta, que describe otros recursos. Los registros de catálogo para materiales bibliotecarios son un ejemplo común de metadatos. Mientras que los recursos son interesantes para el usuario final, los metadatos son interesantes para las personas o programas que tienen que gestionar la información. Los metadatos ayudan en el proceso de recuperar información para capacitar a los usuarios a descubrir inicialmente la existencia de la información, a localizarla y después a determinar si es la información que el usuario quiere. Usualmente los metadatos describen el contenido, la descripción física, la localización, el tipo y la forma

de la información, y la información necesaria para la gestión, incluyendo la historia de la migración, fechas de expiración, seguridad, autenticación y formatos de fichero. (<http://info.wgbh.org/upf/glossary.html>). 2. Los metadatos son datos estructurados, codificados, que describen características de entidades que contienen información, para ayudar en la identificación, descubrimiento, evaluación, y gestión de las entidades descritas (<http://www.ala.org/alcts/organization/ccs/ccda/tf-meta3.html>).

Más ilustrativa resulta acaso la definición de PREMIS (OCLC/RLG Working Group on Preservation Metadata, 2001):

Quizá más que cualquier otro medio, la información digital requiere metadatos detallados para asegurar su conservación y accesibilidad por generaciones futuras. Los metadatos requeridos para conservar un recurso de información tradicional, como un libro, son completamente claros. Un libro es un ítem físico, estático: sus límites están claramente delineados y su contenido no puede alterarse en el curso del tiempo. Además, es poco probable que la tecnología de acceso a un libro —la inspección visual humana— devenga obsoleta. Así, es probable que los requisitos de metadatos primarios de acceso a un libro sean de naturaleza descriptiva —esto es, que faciliten su recuperación— más que administrativa o estructural. En un entorno digital, asegurar que un objeto de información ‘existe físicamente’ es análogo a conservar su cadena de bits sobre un medio de almacenamiento digital no volátil. Esto, sin embargo, es solo una parte del proceso de conservación. Los objetos digitales no son inmutables: por tanto, debe mantenerse la historia del cambio del objeto en el curso del tiempo para asegurar su autenticidad e integridad. Las tecnologías de acceso para objetos digitales devienen frecuentemente obsoletas: por tanto, puede que sea necesario encapsular junto con el objeto información sobre el entorno de hardware, el sistema operativo y el software de representación relevantes. Toda esta información, así como otras formas de descripción y documentación, pueden capturarse en los metadatos asociados a un objeto digital. Los metadatos de conservación están destinados a soportar y facilitar la retención a largo plazo de la información digital. La National Library of Australia proporciona una visión general de los tipos de información que pueden caer en esta categoría. En particular, los metadatos de conservación pueden utilizarse para: almacenar información técnica que apoye las decisiones y acciones de conservación, documentar las acciones de conservación emprendidas, como las políticas de emigración o emulación; registrar los efectos de las estrategias de conservación; asegurar la autenticidad de los recursos digitales en el curso del tiempo; anotar la información sobre gestión de colecciones y gestión de derechos. Los tipos de información enumerados afrontan dos objetivos funcionales: 1) proporcionar a los gestores de la conservación el conocimiento suficiente para tomar acciones adecuadas para mantener la cadena de bits de un objeto digital a largo plazo; y 2) asegurar que el contenido de un objeto archivado puede ser representado e interpretado, pese a los cambios futuros en las tecnologías de acceso.

La New Zealand National Library, por último, hace suyas las definiciones facilitadas por la National Library of Australia y PREMIS, aunque su enfoque diferirá sustancialmente de ambos proyectos.

Por nuestra parte, no nos interesa añadir una nueva definición de metadatos ni de metadatos de conservación a las muchas existentes. Nos limitaremos a recapitular y extraer aquellas características de los metadatos de conservación, definidas más arriba, que nos parecen merecedoras de ser tomadas en cuenta:

- (1) Así, los metadatos de conservación son aquellos que están unidos mediante asociación o anidamiento con un objeto digital, a efectos de conservar el contenido y la estructura de este objeto.
- (2) Si los metadatos de conservación se solapan con otras series o tipos de metadatos es cuestión irrelevante, en la medida en que los metadatos son datos sobre datos en un determinado contexto. Diferentes contextos requerirán probablemente el uso de los mismos metadatos “con otra acepción”. Es preciso, no obstante, que estas distintas acepciones siempre queden claras al usuario y se hagan explícitas en la documentación asociada a la serie de metadatos.
- (3) Además, los metadatos de conservación deben servir para que, a largo plazo, los contenidos de los objetos digitales puedan ser representados e interpretados.
- (4) Entendemos que un objeto es básicamente una cadena de bits, y es sobre esta cadena sobre la que tienen que actuar los metadatos. Otras definiciones de objetos pueden ser adecuadas en otros contextos.
- (5) Los peligros que los metadatos de conservación pretenden minimizar son, sobre todo, los derivados de la volatilidad de los soportes y de la obsolescencia de los entornos. Existen otros aspectos que deben documentarse — gestión de derechos, historia del cambio, etc. —, pero en filigrana los metadatos de conservación deben proporcionar información de tipo tecnológico.
- (6) Consideraremos que una serie de metadatos de conservación cumple en líneas generales su cometido si proporciona información suficiente acerca de la procedencia, la referencia, la autenticidad y el contexto.

De acuerdo con todo ello, entenderemos por metadatos de conservación la infraestructura de información que apoya los procesos asociados con la conservación digital. De manera más específica, es la información necesaria para mantener la viabilidad, la representación y la comprensión de recursos digitales a largo plazo. La viabilidad requiere que la cadena de bits de un objeto digital archivado permanezca intacta y legible a partir de los medios digitales sobre los que se ha almacenado. La representación se refiere a la traducción de la cadena de bits a una forma que pueda ser vista por usuarios humanos, o procesada por ordenadores. La comprensión implica el proporcionar suficiente información para que el contenido representado pueda ser interpretado y comprendido por sus usuarios

pretendidos. Los metadatos de conservación pueden servir como entrada a los procesos de conservación, y también registrar la salida de estos mismos procesos (OCLC/RLG Working Group on Preservation Metadata, 2002, p. 1).

3. Reference Model for an Open Archival Information System (OAIS)

El modelo de referencia OAIS es resultado de una comisión de la Organización Internacional de Normalización al Consultative Committee for Space Data Systems, para la elaboración de una recomendación que habría de especificar los requisitos y responsabilidades a cumplir por un sistema de información archivística abierto, con el compromiso de conservar materiales a largo plazo para una comunidad definida (Consultative Committee for Space Data Systems, 2002, p. 1-1). El modelo OAIS define *archivo* de una manera lo suficientemente amplia como para poder ser adoptado en diversidad de contextos, incluidos los archivos convencionales. De acuerdo con OAIS, un archivo es “una organización que pretende conservar información para su acceso y uso por una comunidad determinada” (*ibidem*, p. 1-8). Se trata tanto de un modelo funcional como de un modelo de información. En el marco del presente trabajo, interesa especialmente su modelo informativo, aunque no se descuidan las principales líneas funcionales. Por otra parte, aunque OAIS afronta todas las facetas relacionadas con la gestión de un archivo digital, existe una fuerte orientación hacia las cuestiones referentes a la conservación a largo plazo. OAIS, además, se define como un sistema abierto y, por tanto, independiente de aplicaciones específicas. Finalmente, de igual modo que en el documento base del modelo de referencia, se utilizan con el mismo significado los términos *OAIS* y *archivo*.

Contextualmente, en un OAIS participan los siguientes actores: productor, gestión, consumidor y, en el centro de todos ellos, el OAIS (archivo) (*ibidem*, p. 2-2).

- (1) El productor son las personas o sistemas que proporcionan la información que ha de conservarse.
- (2) La gestión se identifica con las funciones propias de la alta gestión.
- (3) El consumidor son las personas o sistemas que interactúan con el OAIS para obtener de él información considerada de interés. Un tipo especial de consumidor es la comunidad determinada para la que se conserva la información, y que debiera ser capaz de comprender esta (*ibidem*, p. 2-2).

En el entorno OAIS, la información es “cualquier tipo de conocimiento que puede ser intercambiado, y esta información siempre se expresa (por ejemplo, se representa), mediante algún tipo de datos” (*ibidem*, p. 2-3). En un contexto digital, esta información viene dada típicamente por una cadena de bits, que son un objeto de datos. No obstante, en un archivo pueden coexistir objetos digitales —compuestos de una serie de secuencias de bits— y objetos físicos —un objeto

con propiedades físicamente observables. A diferencia de otros modelos, OAIS acentúa el carácter informático —cadena de bits—, como básico del objeto digital. Para que el objeto de datos sea correctamente interpretado por los usuarios, se precisan además otros dos componentes:

- (1) La *base de conocimientos del usuario*, o el conjunto de referencias que posee este y que le permiten interpretar la información en su contexto. Por ejemplo, a un usuario español se le supone una gramática de la lengua española incorporada a su base de conocimiento.
- (2) La *información sobre la representación*, que es información externa a la base de conocimiento del usuario, y que la complementa a fin de facilitar la correcta interpretación de la información o del objeto de datos. Por ejemplo, si un usuario español accede a un documento en inglés, necesitará información sobre la representación, en forma de diccionarios, gramáticas, etc., que no forman parte de su base de conocimiento. La información sobre la representación es también un objeto, que necesita ser interpretado mediante otra secuencia de información sobre la representación, etc. La información sobre la representación, además, puede ser estructural —por ejemplo, la que nos permite entender el objeto como una secuencia de caracteres ASCII— o semántica —por ejemplo, la que nos permite entender el objeto como un documento en español.

En el entorno OAIS, el objeto de datos más la representación de la información necesaria para interpretarlo constituyen un objeto de información. El objeto de información contiene información estructural e información sobre el contenido. La información sobre el contenido es el objetivo principal de las tareas de conservación, y está constituida por el objeto de datos de contenido y la representación de la información asociada. Una vez que se defina determinada información sobre el contenido, es posible definir también la información sobre descripción de la conservación, que “se aplica a la información sobre el contenido y es necesaria para conservar la información sobre el contenido, para asegurar que se identifica claramente, y para comprender el entorno en el que se creó la información sobre el contenido” (*ibidem*, p. 2-6). Ambas, la información sobre el contenido y la información sobre descripción de la conservación, constituyen un paquete de información. Este paquete de información se puede recuperar mediante información, externa a él, llamada *información descriptiva*. Finalmente la información sobre empaquetamiento es aquella que permite agrupar la información sobre el contenido y la información sobre descripción de la conservación. Así, pues, un paquete de información consta de cuatro objetos de información:

- (1) Información sobre el *contenido*, que es “un objeto de datos de contenido junto con su información sobre la representación” (*ibidem*, p. 4-24).

- (2) Información sobre la descripción de la *conservación*, que “permite la comprensión de la información de contenido durante un período indefinido de tiempo” (*ibidem*, p. 4-27).
- (3) Información de *empaquetamiento*, que es “la información que, real o lógicamente, agrupa o relaciona los componentes de un paquete en una entidad identificable sobre un medio específico” (*ibidem*, p. 4-29).
- (4) Información *descriptiva*, “que contiene los datos que sirven como entrada de los documentos o aplicaciones llamados ayudas al acceso” (*ibidem*, p. 4-30).

Por su parte, la información sobre descripción de la conservación puede ser, a su vez, de cuatro tipos:

- (1) *Procedencia*: describe la fuente de la información sobre el contenido, quién ha tenido su custodia desde su origen y su historia (incluida la historia del procesamiento).
- (2) *Contexto*: describe cómo se relaciona la información sobre el contenido con otra información externa al paquete de información. Por ejemplo, describiría por qué se produjo la información sobre el contenido, y puede incluir una descripción de cómo se relaciona con otro objeto de información disponible sobre el contenido.
- (3) *Referencia*: proporciona uno o más identificadores, o sistemas de identificadores, mediante los cuales la información sobre el contenido puede identificarse de manera única.
- (4) *Seguridad*: proporciona un envoltorio, o una concha protectora, que protege la información sobre el contenido de alteraciones no documentadas. Por ejemplo, puede implicar un contador de la información sobre el contenido de un paquete de información digital (*ibidem*, p. 2-6).

Como reflejo de la existencia de diferentes actores en el proceso archivístico, existen variantes de paquetes de información:

- (1) El paquete de información sobre *sumisión* (SIP) es el enviado por el productor al archivo.
- (2) El paquete de información *archivística* (AIP) es el conservado a efectos de conservación, y el que nos interesa en el presente contexto.
- (3) El paquete de información sobre la *diseminación* (DIP) es el que recibe el consumidor en respuesta a una demanda al archivo.

Añadamos, finalmente que tanto el modelo funcional como el informativo de OAIS se representan mediante una versión simplificada de Unified Modeling Language (UML), lo cual contribuye notablemente a la formalización de su gramática.

4. Conceptos sobre los que se trabaja

De acuerdo con este modelo de referencia, en el presente trabajo se tomará en consideración el paquete de información archivística, y, de manera específica, los metadatos de conservación. En concreto, de los cinco proyectos mencionados, tres de ellos están explícitamente basados en el modelo de referencia OAIS, y dos son independientes, aunque afirman su conocimiento del citado modelo de referencia. Hemos excluido proyectos de interés, como PADI o The Joint Information Systems Committee (JISC), porque, aunque constituyen inestimables fuentes de recursos relativos a digitalización y metadatos de conservación, no proponen series de este tipo de metadatos.

Para la evaluación de los proyectos que siguen hemos utilizado los conocidos principios de modularidad, extensibilidad, refinamiento y multilingüismo, expuestos en el bien aceptado texto *Metadata: Principles and Practicalities*. A estos principios hemos añadido como criterios de valoración los también conocidos conceptos de interoperabilidad y reusabilidad. Por último, hemos tomado en consideración la existencia de una gramática formal y de una gramática formalizada, entendiéndose por la primera un lenguaje dotado explícitamente de una semántica y una sintaxis, y por la segunda la traslación de ese lenguaje a una notación lógica, por ejemplo XML. De acuerdo con el citado texto, definimos estos principios de la siguiente manera:

- *Modularidad*: “La modularidad de los metadatos [...] permite a los diseñadores de esquemas de metadatos crear nuevos conjuntos basados en esquemas de metadatos establecidos y beneficiarse de las mejores prácticas observadas, mejor que reinventar elementos de nuevo” (Duval et al., 2002).
- *Extensibilidad*: “Los sistemas de metadatos deben permitir extensiones para poder acomodar las necesidades particulares de una aplicación dada. Algunos elementos de metadatos es probable que se encuentren en muchos esquemas de metadatos [...] otros serán específicos de aplicaciones o dominios particulares. Las arquitecturas de metadatos deben acomodar fácilmente la noción de un esquema base con elementos adicionales que ajusten una aplicación dada a necesidades locales o específicas de dominio sin comprometer indebidamente la interoperabilidad proporcionada por el esquema base. Otra aplicación que encuentre tales extensiones debiera ser capaz de ignorarlas, aun haciendo uso de los elementos comprendidos por ambas” (*ibidem*).
- *Refinamiento*: “Los dominios de aplicación diferirán según el grado de detalle que es necesario o deseable. El diseño de normas de metadatos debiera permitir a los diseñadores de esquemas elegir un nivel de detalle adecuado a una aplicación dada. Poblar bases de datos con metadatos es

costoso, de modo que existen incentivos económicos para crear metadatos con el detalle suficiente para satisfacer los requisitos funcionales de una aplicación, pero no más” (*ibidem*).

- *Multilingüismo*: “Es esencial adoptar arquitecturas de metadatos que respeten la diversidad lingüística y cultural. La Web, en cuanto sistema de información global, es importante por el hecho de que afronta un acceso sin precedentes a recursos de alcance global. Sin embargo, a menos que tales recursos puedan estar disponibles para los usuarios en sus lenguas nativas, en series de caracteres adecuados, y con metadatos apropiados a la gestión de recursos, la Web no logrará su potencial como sistema de información global. Las normas tratan típicamente estas cuestiones mediante los procesos complementarios de internacionalización y localización: el primer proceso se refiere a la creación de normas ‘neutrales’, mientras que el último se refiere a la adaptación de tales normas neutrales a un contexto local” (*ibidem*).

A estos principios fundamentales añadimos la definición de los otros dos conceptos, relacionados de manera más estrecha con los entornos informáticos:

- *Interoperabilidad*, o “la capacidad para comunicar del software y hardware sobre múltiples máquinas de múltiples vendedores” (FOLDOC, 1993-2006).
- *Reusabilidad* (*reuse* en el original), o “utilización de código desarrollado para una aplicación en otra aplicación” (*ibidem*).

5. Los metadatos para la conservación

5.1. Preservation Metadata for Digital Collections de la National Library of Australia (NLA)

El proyecto de la NLA no sigue el modelo de referencia OAIS, aunque reconoce su conocimiento y uso a efectos de consulta. Su modelo de metadatos de conservación se define explícitamente como “de salida”, no “de entrada”; es decir, no interesa tanto saber qué metadatos se introducirán como saber los metadatos que pueden extraerse del sistema. Define tres niveles de granularidad: colección, objeto y fichero; menciona la cardinalidad de los metadatos solo en términos de repetible/no repetible, y rechaza explícitamente la declaración de obligatoriedad de los elementos, reducida a esencial/deseable. Por otra parte, intenta ser más exhaustivo que otros proyectos, aunque este hecho le confiere cierto carácter de provisionalidad, enunciado en su estatuto de borrador y en los comentarios acerca de la oportunidad y/o solapamiento de algunos metadatos. Como hará posteriormente el modelo neozelandés con mayor sofisticación, el proyecto de la NLA discrimina “obra” —identificada con el concepto— y “manifestación” —identificada con la instancia física de aquella. La serie de metadatos

de la NLA tiene veinticinco componentes de primer nivel, descomponiéndose el componente “descripción de fichero” en distintos tipos de archivo y estos, a su vez, en las características descriptivas relevantes para cada tipo.

5.2. Networked European Deposit Library (NEDLIB)

El proyecto NEDLIB adopta el modelo de referencia OAIS, aunque desde una perspectiva diferente a la expuesta aquí, y de manera en cierto modo más exhaustiva, a pesar de que su serie final de metadatos es menor. Indiquemos, ante todo, que la finalidad última de NEDLIB no es proporcionar una serie de metadatos de conservación, sino un modelo de biblioteca de depósito digital. Los metadatos de conservación son solo una parte del proyecto. En primer lugar, NEDLIB desarrolla su concepto de flujo de trabajos de un sistema de depósito de publicaciones electrónicas (DSEP) a partir del modelo funcional, no del modelo de información, de OAIS. De este modo, prácticamente todas las funciones de un OAIS requieren la asignación de metadatos de un tipo u otro. En segundo lugar, la serie de metadatos de conservación propuesta por NEDLIB parte de la asunción de que, a efectos de conservación, la información que se precisa es básicamente técnica. Metodológicamente, discrimina los metadatos necesarios para la información de representación, y aquellos necesarios para la información de descripción de la conservación. Para elaborar el primer conjunto, extiende el modelo multicapa expuesto en el Anexo E de OAIS, definiendo cinco capas de carácter técnico —física, binaria, estructural, de objeto y de aplicación—, para las que resulta preciso incorporar metadatos. Por lo demás, NEDLIB únicamente define un cuerpo central de metadatos considerados esenciales para la gestión de la conservación. No hace mención a la granularidad, ni a niveles de agregación ni de detalle en los que estos metadatos debieran utilizarse, y expresa la cardinalidad únicamente en términos de repetible/no repetible. Define la obligatoriedad/no obligatoriedad de los metadatos. NEDLIB propone cuarenta y ocho elementos de metadatos, con hasta cuatro niveles de anidamiento, y divididos en dos bloques padre: metadatos para la representación de la información y metadatos para la información de conservación y descripción.

5.3. CURL Exemplars in Digital Archives (CEDARS)

CEDARS es un proyecto conjunto de eLIB (The Electronic Libraries Programme), el ya citado JISC y el CURL (Consortium of University Research Libraries). Se apoya de igual modo en un análisis del modelo de referencia OAIS, en este caso de su modelo de información, de forma que reduce el énfasis en los metadatos técnicos y extrae un conjunto de metadatos, explícitamente declarados como de alto nivel —lo que en la práctica viene a significar altamente generales— y orientados a describir la información de representación y la información de descripción de la conservación. El resultado es un conjunto de metadatos que

documentan aspectos indudablemente relacionados con la conservación (gestión de derechos, historia de la gestión, etc.), pero que han sido parcial o totalmente asumidos por otros modelos de metadatos, y resultan en cierto modo insuficientes para la conservación técnica. CEDARS propone una estructura de metadatos, en la que no especifica granularidad ni cardinalidad de los mismos. Sin embargo, ha desarrollado su trabajo en otros sentidos más positivos, como la elaboración de un prototipo de archivo distribuido. La estructura de metadatos CEDARS propone cuarenta elementos bajo el elemento contenedor “paquete de información”. Los elementos pueden estar anidados hasta un nivel siete.

5.4. PREMIS (PREservation Metadata: Implementation Strategies)

PREMIS constituye un esfuerzo conjunto OCLC/RLG, entre cuyos miembros se encuentran también representantes de otros grupos como UKOLN (UK Office for Library Networking), NLA, CEDARS y NEDLIB. Una de las primeras tareas del grupo de trabajo fue una revisión del estado del arte, a partir del modelo de referencia OAIS, y tomando como objetos de un detallado análisis comparativo los proyectos NEDLIB, CEDARS y NLA, además de las series de elementos de los Digital Repository Services de la Universidad de Harvard, que no siguen el modelo OAIS y no son de acceso público. En una segunda fase, se abandonó el modelo de Harvard y se retomó el trabajo previo de OCLC —OCLC Digital Archive Service— para definir un conjunto mínimo de metadatos. En la segunda parte del proyecto, y a partir de los descubrimientos anteriores, se propuso una estructura de metadatos que atendía tanto al modelo de referencia OAIS como a los resultados de esfuerzos precedentes. El proyecto siguió el modelo de información de un OAIS, poniendo el énfasis en la asignación de metadatos de conservación para la información de contenido y la información de descripción de la conservación, estableciendo dos bloques de recomendaciones y profundizando en el análisis de los tipos de información, de tal modo que se puedan definir metadatos con un mayor grado de detalle.

Durante la fase de elaboración del proyecto, y siempre que resultó posible, PREMIS extrajo sus elementos de metadatos de proyectos anteriores; por lo demás, no hizo mención a la granularidad en el uso de metadatos, la cardinalidad de estos ni los niveles de agregación aplicables. Ello se debía a la decisión de dejar un alto grado de autonomía a los diferentes depósitos, que debieran tener distintas necesidades con relación a aquellas características. Del mismo modo, aunque teóricamente los niveles estructurales de la propuesta tendrían que considerarse como contenedores de los metadatos, lo cierto es que tales niveles también funcionaban como metadatos, dependiendo de las necesidades de granularidad de los depósitos. Al contrario, la necesidad de descender a un grado detallado de granularidad implicaba que dentro de la estructura pueden anidarse

elementos procedentes de series externas. La propuesta PREMIS fue, de acuerdo con esto, y junto con la de NZNL, la más compleja, en la medida en que, además de un cuerpo de metadatos con varios niveles de agregación, reconocía varios procesos, cada uno de los cuales podía anidar eventos, que a su vez podían anidar elementos de metadatos.

En la actualidad el proyecto PREMIS se encuentra bien consolidado, y ha generado varias herramientas que incluyen un diccionario de datos y varios schemas XML. De igual modo, la versión final incluye cuatro entidades sobre las que se aplican metadatos: objetos, eventos, agentes y derechos. Para las cuatro entidades se han definido schemas, así como para el contenedor PREMIS.

5.5. Metadata Standards Framework-Preservation Metadata, de la New Zealand National Library

El proyecto de la New Zealand National Library —con una primera versión en 2002, revisada en 2003— reconoce las referencias a proyectos anteriores, especialmente PREMIS, así como su conocimiento del modelo de referencia OAIS. De hecho, incorpora referencias cruzadas entre su serie de metadatos y estas otras experiencias. Sin embargo, su referente inmediato es el proyecto de la NLA, sobre el que la NZNL establece un buen número de refinamientos. El proyecto tiene como objeto principal de conservación el llamado “máster de conservación”, entendido como la representación del “mejor esfuerzo de creación de un objeto de trabajo, a partir de alguna forma de original, tal como se suministró a o fue adquirido por la biblioteca [...] será una representación de ese original en un objeto que puede ser conservado, gestionado y diseminado a lo largo del tiempo” (New Zealand National Library, 2003, p. 6). Por lo demás, un master de conservación es dinámico, en la medida en que puede quedar obsoleto, sufrir procesos de conversión o migración, etc.

Por otra parte, el proyecto define varios niveles de agregación organizados en lo que podría ser un par de coordenadas: por una parte, las entidades —procesos, objetos, ficheros y modificaciones de metadatos—, sujetas a diversas relaciones entre ellas; por otra, los tipos de objetos digitales —objeto simple, objeto complejo y grupo de objetos. Aunque cada serie de metadatos de conservación se aplica a un solo objeto lógico, este objeto puede estar compuesto de diferentes maneras. En el proyecto de la NZNL, cada entidad tiene su propio conjunto de metadatos, de los que no se menciona la cardinalidad. Con respecto a la obligatoriedad, se describe, a semejanza de la NLA, en términos de esencial/no esencial. La NZNL lleva a cabo un análisis más exhaustivo, al discriminar entidades o, por ejemplo, al contemplar posibles solapamientos con otras series de metadatos. De igual modo, el hecho de estar integrado en un proyecto más amplio le proporciona el soporte de un cuidado análisis de normas existentes.

6. Análisis de las series de metadatos

En primer lugar se elaboró una tabla comparativa de las distintas series. Para la elaboración de la tabla se tomó como referencia el modelo básico RLG/OCLC, en la medida en que se trata de un proyecto integrador. Aquellos elementos extraídos de otras series de metadatos se indicaron con una X. Los elementos de otras series que no fueron incluidos en el trabajo RLG/OCLC fueron analizados para anidarlos en el punto de referencia potencialmente adecuado. Obviamente, no todos los elementos de todos los proyectos fueron incluidos, y, en la tabla, existía un cierto grado de duplicación y solapamiento. No obstante, resultó útil a efectos de análisis y depuración. Al proyecto de la NZNL se le dio un tratamiento escueto, puesto que era posterior al proyecto PREMIS y se distanciaba conceptualmente de él. Sin embargo, fue de utilidad en un momento posterior del análisis. Se incorporó, como se dijo, el contraste con la primera serie de metadatos del OCLC, puesto que algunos elementos de PREMIS se extrajeron de ella. Debe tenerse en cuenta que esta tabla es altamente refinable. Por ejemplo, algunos de los elementos que aparecen en ella —como los tipos estructurales procedentes de la NLA— podrían pasar a ser atributos satisfechos mediante tipos MIME. De otros que se repiten en distintos lugares —documentación, localización— debiera definirse su cardinalidad. No se han definido atributos para el modelo de origen, el nivel de objeto de acuerdo con la NLA, o el tipo de entidad según la NZNL.

Los elementos expuestos en la citada tabla se tradujeron a una instancia de documento XML en bruto, es decir, sin análisis para determinar duplicaciones, anidamientos, etc. Tanto la tabla como la instancia de documento deben considerarse herramientas de trabajo, para un análisis posterior en el que se definan los elementos realmente útiles, se eliminen los duplicados, se asignen atributos, etc. A efectos de claridad, se incluyó información extraída de las definiciones y/o los ejemplos de las series de metadatos implicadas. La corrección de la formación de la instancia de documento fue chequeada con Microsoft Internet Explorer 6.0, XMLSpy, Stylus Studio y Topologi 2.0. El siguiente paso, pues, fue generar una gramática formalizada de este conjunto de metadatos de conservación. Para la generación automática de un schema XML se utilizaron el mecanismo Microsoft XSD Inference 1.0 y Stylus Studio, siendo los resultados diferentes en uno y otro caso. Para la generación de una DTD se utilizó Stylus Studio. Se verificó que los schemas y la DTD eran correctos mediante XMLSpy, una herramienta distinta de aquellas con las que se generaron.

Para representar la serie de metadatos en un entorno de base de datos relacional se seleccionó el formato MARC, debido a su extendida utilización, y la serie de metadatos PREMIS, con el fin de elaborar un prototipo. El aplicativo utilizado fue Archivo 3000. En este caso, y a diferencia de lo que sucede con XML, los datos se definieron posicionalmente, de tal modo que solo se permitían dos niveles de ani-

damiento: etiqueta y subcampo. Se solventó esta limitación definiendo, mediante campos locales, los metadatos de nivel superior como etiquetas y subcampos, e incorporándolos a los campos de información descriptiva. Los metadatos de nivel inferior se incorporaron mediante la definición de menús contextuales que permitían seleccionar de un rango de metadatos los procedentes en cada campo, y generar la correspondiente etiqueta XML, en la que se introducía la información. Alternativamente, y para aquellos casos en los que no se deseara combinar los metadatos de conservación con la información descriptiva, se generó un solo campo, en el que era posible insertar la plantilla de la instancia de documento XML y trabajar sobre ella mediante editor de texto incluido en el aplicativo.

Por último, y también con el objeto de evitar la mezcla de información descriptiva y de conservación, se permitió que, en lugar de anidar los metadatos de conservación en la descripción, se asociara esta tanto con su objeto digital como con la instancia de documento XML donde se recogen estos metadatos de conservación.

7. Conclusiones

Si aceptamos que una de las funciones del archivo —incluidos los archivos en entornos digitales, conceptualmente mucho más amplios— es la conservación a largo plazo de los documentos para beneficio de una determinada comunidad de usuarios futura, y si aceptamos también que el entorno tecnológico en el que nos desenvolvemos es cada vez más complejo y cambiante, sin que podamos definir líneas maestras más allá de un medio plazo razonable, entonces debemos admitir que la incorporación a los objetos bajo nuestra custodia de información que permita su representación a lo largo del tiempo y a través de los cambios de tecnología es una más de las obligaciones archivísticas. Afortunadamente, en este momento disponemos de modelos de referencia y de metadatos, cuyo uso y satisfacción no siempre son sencillos, pero cuya manejabilidad es susceptible de optimización, acudiendo tanto a tecnologías implantadas como a tecnologías emergentes, o a una o varias combinaciones de estas.

La propuesta metodológica expuesta en el presente trabajo, y los trabajos prácticos que la validaron, debieran servir para seguir profundizando en la tarea de garantizar el acceso al patrimonio digital a las generaciones futuras.

Notas

- (1) Poco después de que el trabajo hubiera concluido, el modelo de referencia OAIS fue adoptado como norma ISO: ISO 14721: 2003: Space data and information transfer systems - Open archival information system - Reference model (2003). Geneva: International Organization for Standardization, 2003.

- (2) El presente texto no reproduce los trabajos anteriores, simplemente los menciona, por motivos de espacio. Puede accederse a estos documentos de trabajo contactando con el autor del presente texto.

Referencias

- Borbinha, José Luis; Cardoso, Fernando; Freire, Nuno (2000). Nedlib Glossary. La Haya: Nedlib, 2000. URL: <<http://nedlib.kb.nl/glossary.pdf>>. Consultado: 2003-12-31.
- Consultative Committee for Space Data Systems (2002). Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-B-1: Blue Book. Washington DC: CCSDS Secretariat, 2002. URL: <<http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>>. Consultado: 2003-12-31.
- Cedars: CURL Exemplars in Digital Archives (2002). Leeds: CEDARS, 2002. URL: <<http://www.leeds.ac.uk/cedars/guideto/metadata/>>. Consultado: 2004-01-01.
- Duval, Erik; Hodgins, Wayne; Sutton, Stuart; Weibel, Stuart L. (2002). Metadata Principles and Practicalities. // D-Lib Magazine. 8:4 (April 2002). URL: <<http://www.dlib.org/dlib/april02/weibel/04weibel.html>>. Consultado: 2004-05-16.
- FOLDOC: Free On Line Dictionary of Computing (1993). Londres: Imperial College Department of Computing, 1993-2006. URL: <<http://foldoc.org/>>. Consultado: 2006-08-19.
- ISO 14721: 2003: Space data and information transfer systems - Open archival information system - Reference model (2003). Ginebra: International Organization for Standardization, 2003.
- ISO/IEC 11179-1:2004: Information technology - Metadata registries (MDR) - Part 1: Framework (2004). Ginebra: International Organization for Standardization, 2003.
- Lupovici, Catherine, Masanès, Julien (2002). Metadata for long term-preservation. La Haya: Nedlib, 2000. URL: <<http://nedlib.kb.nl/results/NEDLIBmetadata.pdf>>. Consultado: 2006-08-19.
- PADI: Preserving Access to Digital Information (2003). Canberra: National Library of Australia, 2003. URL: <<http://www.nla.gov.au/padi/topics/32.html>>. Consultado: 2004-05-18.
- New Zealand National Library (2003). Metadata Standards Framework: Preservation Metadata: Revised. Wellington: New Zealand National Library, 2003. URL: <http://www.natlib.govt.nz/files/4initiatives_metaschema_revised.pdf>. Consultado: 2004-01-01.
- OCLC/RLG Working Group on Preservation Metadata (2002). Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects: A Report by The OCLC/RLG Working Group on Preservation Metadata. Dublín: Online Computer Library Center; Mountain View: Research Libraries Group, 2002. URL: <<http://www.oclc.org/research/pmwg/>>. Consultado: 2003-12-31.
- OCLC/RLG Working Group on Preservation Metadata (2001). Preservation Metadata for Digital Objects: A Review of the State of the Art. Dublín: Online Computer Library

- Center; Mountain View: Research Libraries Group, 2001. URL: <http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf>. Consultado: 2006-08-19.
- Phillips, Margaret; Woodyard, Deborah; Bradley, Kevin; Webb, Colin (1999). Preservation Metadata for Digital Collections. Canberra: National Library of Australia, 1999. URL: <<http://www.nla.gov.au/preserve/pmeta.html>>. Consultado: 2004-01-01.
- PREMIS Working Group (2005). Data Dictionary for Preservation Metadata. Dublín: Online Computer Library Center; Mountain View: Research Libraries Group, 2005. URL: <<http://www.oclc.org/research/projects/pmwg/premis-final.pdf>>. Consultado: 2006-08-19.
- PREMIS Working Group (2006). PREMIS: PREservation Metadata Implementation Strategies. Dublín: Online Computer Library Center; Mountain View: Research Libraries Group, 2006. URL: <<http://www.oclc.org/research/projects/pmwg/>>. Consultado: 2006-08-19.
- PREMIS Working Group (2006b). XML Schemas for the PREMIS Data Dictionary. Dublín: Online Computer Library Center; Mountain View: Research Libraries Group, 2006. URL: <<http://www.loc.gov/standards/premis/schemas.html>>. Consultado: 2006-08-19.
- The Joint Information Systems Committee (2003). Londres: JISC, 2003-2006. URL: <<http://www.jisc.ac.uk/index.cfm?name=home>>. Consultado: 2004-01-01.
- Van der Werf, Titia (2000). The Deposit System for Electronic Publications: a Process Model. La Haya: Nedlib, 2000.