

Introducción a los

Metadatos

vías a la información digital



PROPERTY OF KAREN SCHMIDT

Introducción a los

Metadatos

vías a la información digital

editado por Murtha Baca

traducido al español por Marisol Jacas-Santoll



Imagen de la portada: Detalle de *El encuentro de los tres reyes, con David e Isaías*, anterior a 1480, obra del Maestro del retablo de San Bartolomé, J. Paul Getty Museum, Los Angeles, California. Oleo y hoja de oro sobre panel, 24³/₄ × 28¹/₆ pulgadas (68.8 × 71.2 cm).

© 1999 The J. Paul Getty Trust
Impreso en los Estados Unidos de
América

Library of Congress Cataloging-in-
Publication Data

Introduction to metadata. Spanish
Introducción a los metadatos : vías a
la información digital / editado
por Murtha Baca ; traducido al
español por Marisol Jacas-Santoli.

p. cm.

ISBN 0-89236-535-8

(alk. paper)

1. Database management.
2. Metadata. 3. World Wide Web
(Information retrieval system)

I. Baca, Murtha. II. Getty Informa-
tion Institute. III. Title

QA76.9.D3I59918 1998

005.7—dc21

98-45660

CIP

Indice

iv	Introducción
1	La definición de los metadatos <i>Anne J. Gilliland-Swetland</i>
10	Los metadatos y la World Wide Web <i>Tony Gill</i>
21	Correspondencias entre metadatos e interoperabilidad: ¿Qué significa todo esto? <i>Willy Cromwell-Kessler</i>
25	Estándares para metadatos: Comparación
36	Conclusión
37	Glosario
41	Acrónimos usados en el texto y una selección de direcciones de la World Wide Web
43	Colaboradores

Introducción

Esta publicación quiere servir de base a un aspecto importante pero poco comprendido—y aún en estado de gestación—de la era de la información digital: los metadatos.

Han contribuido al texto profesionales dedicados al desarrollo y a la implementación de estándares de información: Anne J. Gilliland-Swetland presenta una perspectiva general en la que muestra los tipos, las funciones, los atributos y las características de los metadatos, con ejemplos sacados del “mundo real”. Su artículo trata de disipar mitos comúnmente aceptados y de demostrar la importancia y el papel de los metadatos en el universo de la información. El artículo de Tony Gill examina el papel de los metadatos en el contexto de la World Wide Web y analiza dos estándares importantes para metadatos de la Web. Finalmente, Willy Cromwell-Kessler discute cómo la comparación entre distintos estándares para metadatos facilita la interoperabilidad de los sistemas, señalando al mismo tiempo algunos problemas relacionados con este enfoque, así como las ventajas que ofrece y los pasos que habrá que tomar en el futuro.

Se incluye también una tabla en la que se asocian nueve estándares para metadatos relacionados con la información sobre el patrimonio cultural y con el descubrimiento de recursos en la Web. Un glosario y una lista de acrónimos y direcciones de la Web complementan esta “guía” para exploradores del mundo de la información digital.

Los metadatos, particularmente en la World Wide Web, se encuentran aún en sus inicios y no cabe duda de que continuarán desarrollándose en el futuro. Creemos, sin embargo, que, aún en este momento relativamente prematuro, esta cuestión es suficientemente importante como para justificar una publicación. Esperamos que este texto ayude a evitar errores y trabajo en vano a quienes tengan un interés en el debate (ya sean bibliotecarios, especialistas de museos o simplemente individuos que quieran dar información a través de una red electrónica de comunicaciones), para que así puedan tomar decisiones más aptas sobre la información que traten de divulgar o registrar.

La definición de los metadatos

Anne J. Gilliland-Swetland
Department of Library and Information Science,
University of California, Los Angeles

La palabra *metadatos*, literalmente “datos sobre datos”, se ha convertido en un término omnipresente pero que aún ahora tiene significados distintos según los profesionales o especialistas que la usen. Es indiscutible que a medida que la era de la información se convierte en una realidad, será esencial que comprendamos el papel crítico que los distintos tipos de metadatos jugarán en el desarrollo de sistemas de información efectivos, acreditados, flexibles y rigurosos.

Tradicionalmente, los especialistas en información sobre el patrimonio cultural, tales como administradores de museos, bibliotecarios y archiveros, han usado el término *metadatos* con referencia a datos sobre indexación y catalogación creados por ellos mismos para ordenar y, en general, hacer más accesible esa información. Desde los años 60, las bibliotecas, con la ayuda de normas de catalogación y estándares de estructura y de contenido internacionalmente reconocidos tales como MARC (MACHINE-Readable Cataloging format) y encabezamientos temáticos como LCSH (*Library of Congress Subject Headings*), han compartido metadatos descriptivos usando sistemas informatizados; un ejemplo son los catálogos de acceso público en línea. Sin embargo, los métodos y la teoría descriptiva varían considerablemente según los objetivos de los distintos grupos y organizaciones que hacen uso de los metadatos; además, están surgiendo aun otros estándares cuya finalidad es captar las similitudes y, al mismo tiempo, sacar provecho de las diferencias que existen entre dichos grupos u organizaciones. Verdaderamente, desde finales del siglo XIX, los especialistas no habían tenido a su alcance una selección tan fascinante—y a la vez tan desconcertante—de esquemas descriptivos y organizativos.

No obstante, los metadatos son algo más que elementos descriptivos. Hace falta una conceptualización más inclusiva de los metadatos que sea capaz de incorporar nuevos ámbitos dentro de los sistemas digitales de información. Por ejemplo, los depositarios de obras de arte crean metadatos relacionados con la administración, el acceso, la conservación y

el uso de colecciones. También se elaboran metadatos para los registros de adquisición, los catálogos de exhibiciones y los datos sobre el uso, aunque éstos todavía se hagan mayoritariamente en papel. Hoy en día, los sistemas de información integrados, tales como museos virtuales y bibliotecas y archivos digitales, incluyen versiones digitales y descripciones del contenido de las colecciones. La incorporación de otros tipos de metadatos dentro de tales sistemas reafirma su importancia en la gestión y en el mantenimiento de las colecciones. Paul Conway hace referencia a esta cuestión cuando discute el efecto de la digitalización en la conservación de objetos:

El mundo digital transforma los conceptos tradicionales de conservación, de modo que la protección de la integridad física de los objetos se ve relegada por la especificación de la creación y el mantenimiento del objeto, cuya integridad intelectual es su característica primaria.¹

Dado que el usuario puede acceder directamente al contenido a través de una red, se necesitan metadatos para indicar el nombre y el tipo de depositario, para certificar la autenticidad y el contexto del contenido, y para facilitar algunos de los datos que un especialista en información hubiera dado a través de una referencia física o en el marco de una investigación. Por otra parte, los metadatos, cuidadosamente contruidos, pueden proporcionar también puntos de acceso adicionales a contenido del que podrían beneficiarse los sistemas digitales de información.

En dominios de información menos tradicionales, el término metadatos tiene un ámbito de aplicación aun más amplio. Un proveedor de recursos del Internet podría usar metadatos con referencia a información codificada como metaetiquetas HTML con el fin de hacer que un sitio de la Web fuera más fácil de encontrar. Los individuos que digitalizan imágenes puede que identifiquen los metadatos con la información que entran en el encabezamiento de una ficha digital para registrar información sobre una imagen, su proceso de computarización y los derechos asociados con ella. Un archivero de datos de ciencias sociales podría usar el término en referencia a los sistemas y a la documentación necesarios para usar e interpretar una cinta magnética que contuviera datos no procesados. Un archivero de datos electrónicos puede que use el término para designar toda la información contextual, de procesamiento y de uso que es necesaria para identificar y documentar la aplicación, la autenticidad y la integridad de un registro en un sistema electrónico. En todas estas interpretaciones diversas, los metadatos no sólo identifican y describen un

¹ Conway, Paul, *Preservation in the Digital World*, Washington, DC: Commission on Preservation and Access, 1996. <http://www.clir.org/cpa/reports/conway2/>.

objeto informático sino que también documentan su comportamiento, su función y su uso, su relación con otros objetos informáticos y la forma en que debería gestionarse.

Todas estas perspectivas sobre metadatos son importantes en el desarrollo de sistemas de redes de información digital, pero conducen a una concepción de los metadatos demasiado amplia. Para entender mejor esta concepción, es útil dividirla en distintas categorías—metadatos administrativos, descriptivos, de conservación, de uso y, por último, metadatos técnicos—que reflejan aspectos clave de la funcionalidad de los metadatos. La tabla 1 define cada una de estas categorías y da ejemplos de las funciones corrientes que cada una de ellas podría ejecutar en un sistema de información digital.

Además de los distintos tipos de metadatos y de sus diversas funciones, existen varias características asociadas con los mismos. La tabla 2 muestra algunos de los atributos clave de los metadatos con ejemplos.

La creación y la gestión de metadatos se han convertido en una mezcla muy compleja de procesos manuales y automáticos creados por distintos individuos, a través de diversas funciones, durante el lapso de vida

Tabla 1 **Distintos tipos de metadatos y sus funciones**

Tipo	Definición	Ejemplos
Administrativo	Metadatos usados en la gestión y administración de recursos de información	<ul style="list-style-type: none"> • Adquisición de información • Control de derechos y reproducciones • Documentación de requisitos legales • Información sobre localización • Criterios de selección para la digitalización • Control de las distintas versiones
Descriptivo	Metadatos usados para describir o identificar información sobre recursos	<ul style="list-style-type: none"> • Catalogación de registros • Ayudas para la búsqueda • Índices especializados • Relaciones hiperenlazadas entre recursos • Anotaciones de usuarios
Conservación	Metadatos relacionados con la conservación de recursos de información	<ul style="list-style-type: none"> • Documentación de recursos sobre la condición física • Documentación de acciones tomadas para conservar versiones físicas y digitales de los recursos, e.g., regeneración y migración de datos
Técnico	Metadatos relacionados con el funcionamiento de los sistemas o el comportamiento de los metadatos	<ul style="list-style-type: none"> • Documentación de hardware o software • Digitalización de la información, p. ej., formatos, ratios de compresión, ajustes • Control del tiempo de respuesta de los sistemas • Autenticación y seguridad de los datos, p. ej., claves cifradas, contraseñas
Uso	Metadatos relacionados con el nivel y el tipo de uso de los recursos de información	<ul style="list-style-type: none"> • Registro de exhibiciones • Seguimiento de usos y de usuarios • Uso repetido del contenido e información sobre versiones múltiples

Tabla 2 **Atributos y características de los metadatos**

Atributo	Características	Ejemplos
Fuente de los metadatos	<p>Metadatos internos generados para un objeto informático en el momento de su creación o digitalización</p> <p>Metadatos externos relacionados con un objeto informático creados posteriormente, generalmente por alguien distinto del agente original</p>	<ul style="list-style-type: none"> • Nombres de los registros e información sobre la etiqueta inicial • Estructuras de directorio • Formato del registro y esquema de compresión • Fichas de registro y de catalogación • Derechos y otra información legal
Métodos para la creación de metadatos	<p>Metadatos automáticos generados por ordenador</p> <p>Metadatos manuales creados por individuos</p>	<ul style="list-style-type: none"> • Índices de palabras clave • Registros de las operaciones de los usuarios • Copias descriptivas tales como registros de catalogación y metadatos Dublin Core
Carácter de los metadatos	<p>Metadatos creados por individuos que no son ni especialistas temáticos ni en información, generalmente el creador original del objeto informático</p> <p>Metadatos expertos creados por especialistas temáticos o en información, generalmente no se trata del creador original</p>	<ul style="list-style-type: none"> • Metadatos creados para una página Web personal • Sistemas personales de archivo • Encabezamientos temáticos especializados • Registros MARC • Ayudas para la búsqueda en archivos
Estatus	<p>Metadatos estáticos que no cambian nunca una vez creados</p> <p>Metadatos dinámicos que pueden cambiar con el uso o con la manipulación de un objeto informático</p> <p>Metadatos de larga duración para asegurar que el objeto informático siga siendo accesible y se pueda usar</p> <p>Metadatos de corta duración, principalmente de tipo operacional</p>	<ul style="list-style-type: none"> • Título, procedencia y fecha de creación de un recurso de información • Estructura de directorio • Registros de las operaciones de los usuarios • Resolución de imágenes • Formatos técnicos y procesamiento de información • Información sobre derechos • Conservación y administración de la documentación
Estructura	<p>Metadatos estructurados que responden a una estructura previsible, tanto si ésta es estándar como si no lo es</p> <p>Metadatos no estructurados que no responden a una estructura previsible</p>	<ul style="list-style-type: none"> • MARC • TEI y EAD • Formatos de bancos de datos locales • Campos de notas y anotaciones sin estructurar
Semántica	<p>Metadatos controlados que responden a un vocabulario estándar o a un formulario de autoridad</p> <p>Metadatos no controlados que no responden ni a un vocabulario estándar ni a un formulario de autoridad</p>	<ul style="list-style-type: none"> • AAT • ULAN • AACR2 • TGN • Notas de texto libre • Metaetiquetas HTML
Nivel	<p>Metadatos de colección relacionados con colecciones de objetos informáticos</p> <p>Metadatos relacionados con objetos informáticos individuales, a menudo incluidos dentro de una colección</p>	<ul style="list-style-type: none"> • Registro a nivel de colección, p. ej., registro MARC o ayuda para la búsqueda • Índices especializados • Transcripción de fechas y de texto a pie de imagen • Información sobre el formato



Figura 1 **El ciclo de vida de los objetos dentro de un sistema digital**

Creación y versiones múltiples: El objeto entra dentro de un sistema digital por el hecho de haber sido creado digitalmente o bien porque ha sido convertido a un formato digital. Es posible que se produzcan múltiples versiones de ese objeto por razones de conservación, investigación, difusión o incluso para fines relacionados con la creación de un producto. Puede que el agente creador incluya algunos metadatos administrativos y descriptivos.

Organización: Los objetos se organizan manual o automáticamente dentro de la estructura de un sistema digital. Se pueden crear metadatos adicionales para ese objeto por medio de registros, catalogación y procesos de indexación.

Búsqueda y acceso: Los usuarios pueden buscar y acceder a los objetos informáticos almacenados y distribuidos. El sistema de los ordenadores crea metadatos que rastrean los algoritmos de acceso, las operaciones de los usuarios y la efectividad del sistema en cuanto a almacenamiento y extracción de datos.

Utilización: Los objetos hallados se utilizan, se reproducen y se modifican. Se puede crear metadatos relacionados con las anotaciones de los usuarios, el control de los derechos y la verificación de las distintas versiones.

Conservación y disposición: Los objetos informáticos se someten a procesos de regeneración y migración, así como a controles de integridad para asegurar su disponibilidad. Los objetos inactivos o innecesarios pueden ser desechados. Los metadatos pueden documentar tanto la conservación como la eliminación de los objetos.

de un objeto informático.² La figura 1 ilustra las diferentes fases por las que generalmente pasan los objetos informáticos a lo largo de su existencia en un entorno digital.³ A medida que pasan por cada fase, los objetos informáticos adquieren estratos de metadatos con los que pueden asociarse de modos distintos. Los metadatos pueden estar contenidos en el mismo “envoltorio” que el objeto informático, por ejemplo, como información de encabezamiento para el registro de una imagen. Los metadatos pueden estar asociados con el objeto informático a través de indicadores bidireccionales o hiperenlaces. Las relaciones entre los metadatos y los objetos

² Un objeto informático es un elemento o grupo de elementos digitales, independientemente del tipo o del formato, que puede ser interpretado o manipulado como un objeto individual por un ordenador. Este concepto puede ser confuso ya que se puede usar tanto con referencia al contenido real (tal como imágenes digitales) como al contenido sucedáneo (tales como registros de catálogos y ayudas para búsqueda).

³ Modificación del Information Life Cycle, *Social Aspects of Digital Libraries: A Report of the UCLA–NSF Social Aspects of Digital Libraries Workshop*, Los Angeles, CA: Graduate School of Education and Information Studies, November 1996: 7.

informáticos, así como entre los distintos aspectos de los metadatos, pueden también documentarse en un registro de metadatos.⁴ A medida que los diseñadores de sistemas asuman la necesidad de incorporar y gestionar metadatos en los sistemas de información, irán apareciendo nuevos mecanismos para asociar los metadatos con los objetos informáticos.

Aclaración de algunos mitos corrientes relacionados con los metadatos

1. *Los metadatos no tienen que ser necesariamente digitales.* Los especialistas en el patrimonio cultural han estado creando metadatos desde que se empezaron a administrar colecciones. Estos metadatos se irán incorporando a sistemas digitales de información.
2. *Los metadatos no sólo están relacionados con la descripción de un objeto.* Aunque muchos museos, archivos y bibliotecas asocien este término preferentemente con la descripción y la catalogación, los metadatos pueden también referirse al contexto, la gestión, el procesamiento, la conservación y el uso de los recursos descritos.
3. *Los metadatos pueden proceder de distintas fuentes;* puede que los haya producido un individuo (especialista en información o usuario), que los haya creado automáticamente un ordenador o que hayan sido inferidos como consecuencia de la relación con otra fuente, como por ejemplo un hiperenlace.
4. *Los metadatos continúan acumulándose durante el ciclo de vida de un objeto informático.* Los metadatos se crean, se modifican e incluso se desechan en varios momentos a lo largo de la duración de un recurso.

¿Por qué son importantes los metadatos?

Como acabamos de ver en la discusión anterior, los metadatos son un constructo muy complejo que puede resultar costoso tanto de crear como de mantener. ¿Cómo se puede justificar entonces el coste y el esfuerzo que suponen? La evolución de la World Wide Web, así como de otros sistemas digitales de información en red, ha abierto una amplia gama de posibilidades a los especialistas en información, aunque también los haya forzado a enfrentarse a problemas que nunca antes se habían planteado. Sólo el uso

⁴ Joint Workshop on Metadata Registries, Workshop Report, Draft 1.6, 1997.
<http://www.lbl.gov/~olken/EPA/Workshop/report.html>.

de metadatos cuidadosamente elaborados, siguiendo, a ser posible, normas nacionales e internacionales, ha permitido que los especialistas en información pudieran aprovechar algunas de estas nuevas oportunidades.

Más accesibilidad: La efectividad de la búsqueda se puede mejorar significativamente con metadatos detallados y consistentes. Los metadatos también pueden facilitar una búsqueda por múltiples colecciones o bien crear colecciones “virtuales” a partir de materiales distribuidos entre varios depositarios—esto siempre y cuando se trate de los mismos metadatos descriptivos o de metadatos que puedan ser aplicados a través de distintos sitios de la Web. Los sistemas digitales de información, así como los estándares para metadatos que están siendo desarrollados por distintos grupos de especialistas—incorporando elementos comunes tales como la Encoded Archival Description (EAD), el Text Encoding Initiative (TEI) y el Dublin Core—permiten al usuario moverse con facilidad entre las copias descriptivas del objeto informático y las versiones digitales del objeto mismo y, al mismo tiempo, llevar a cabo búsquedas, tanto a nivel del objeto como de la colección, dentro y a través de los sistemas de información.

Retención del contexto: Los museos, los archivos y las bibliotecas no se limitan simplemente a almacenar objetos. En estas instituciones se mantienen objetos que están interrelacionados entre sí y, a la vez, asociados con individuos, lugares, movimientos y acontecimientos. En el mundo digital ocurre a menudo que un objeto de una colección, después de ser digitalizado, se separa tanto de la información con que está catalogado como de sus relaciones con otros objetos de la misma colección. En este sentido, los metadatos juegan un papel de suprema importancia en la documentación y el mantenimiento de estas relaciones, así como en el establecimiento de la autenticidad, la integridad estructural y la exhaustividad de los objetos informáticos. Por ejemplo, cuando se documenta el contenido, el contexto y la estructura de un registro de archivo lo que distingue ese registro de la información descontextualizada son los metadatos de ayuda para la búsqueda.

Uso más amplio: Los sistemas digitales de información para museos y archivos facilitan la difusión de versiones digitales de los objetos a usuarios que, por razones geográficas, económicas o de otra índole, quizá nunca habrían tenido la oportunidad de verlos. Sin embargo, con cada nuevo grupo de usuarios surge la cuestión de cómo acrecentar el acceso intelectual a esos objetos. Los nuevos usuarios puede que tengan necesidades muy distintas de las que tenían los usuarios tradicionales, para los cuales se han diseñado muchos de los servicios de información existentes. Por ejemplo, los maestros y los escolares quizá quieran buscar y usar la información de modo distinto a como lo hacen los investigadores. Los metadatos pueden documentar distintos usos de los sistemas y de su contenido, y esta información puede ser útil a la hora de tomar decisiones

respecto al desarrollo de nuevos sistemas. Los metadatos correctamente estructurados pueden facilitar un número casi infinito de estrategias para buscar información, presentar los resultados e incluso manipular los objetos informáticos sin comprometer su integridad.

Versiones múltiples: La existencia de información y de objetos culturales en forma digital ha incrementado el interés en la posibilidad de crear versiones múltiples y distintas de esos objetos. Este proceso puede que simplemente se limite a la creación de una copia de alta resolución, para efectos de conservación o de investigación, y a una copia de baja resolución que pueda ser transmitida rápidamente a través de una red para tener una referencia instantánea de la imagen. Es posible, también, que se trate de crear una variante o una forma derivada de la imagen para usarla en publicaciones, exhibiciones o escuelas. En cualquier caso, ha de haber metadatos que conecten las distintas versiones y que capten sus similitudes y sus diferencias. Los metadatos tienen también que poder distinguir lo que es cualitativamente diferente entre las distintas versiones digitalizadas de un objeto y la copia impresa original o el objeto mismo.

Cuestiones legales: Los metadatos permiten a los depositarios identificar los numerosos niveles de información sobre los derechos y las restricciones que afectan el uso de un objeto informático y de sus distintas versiones. Los metadatos también documentan otros requisitos legales, o aquellos impuestos sobre los objetos por los donantes—por ejemplo, cuestiones de privacidad o intereses de propiedad.

Conservación: Si los objetos informáticos digitales han de sobrevivir generaciones sucesivas de hardware y software o resistir el cambio a nuevos sistemas, han de contar con metadatos que les permitan existir independientemente del sistema que se use para almacenarlos y encontrarlos. En este sentido, son esenciales los metadatos técnicos, descriptivos y de conservación que sirven para documentar la creación y el mantenimiento de un objeto informático, su conducta y sus relaciones con otros objetos. Cabe notar que para que los objetos informáticos puedan mantener su accesibilidad e inteligibilidad es imprescindible poder conservar y trasladar esos metadatos.

Mejora de sistemas y economía: Los datos técnicos de referencia, la mayoría de los cuales pueden ser recopilados automáticamente por un ordenador, son necesarios para evaluar y perfeccionar los sistemas de modo que sean más efectivos y eficientes desde un punto de vista técnico y económico. Estos datos también pueden usarse en la planificación de nuevos sistemas.

Conclusión

Si los metadatos son rigurosos y consistentes, pueden asistir a los usuarios no tradicionales en una infinidad de nuevas maneras; pueden servir también para crear versiones múltiples de un objeto y para encontrar datos con programas computarizados. Sin embargo, los recursos necesarios y las cuestiones intelectuales y de diseño técnico que comportan no son de poca importancia. Por ejemplo, a medida que van apareciendo nuevos objetos informáticos y nuevos sistemas digitales, los especialistas deben decidir qué aspectos de los metadatos son esenciales para un fin particular o cómo ha de ser de detallado cada nuevo metadato. Probablemente será siempre difícil mantener un equilibrio entre el coste de crear y gestionar metadatos adecuados a las necesidades del momento y el desarrollo de metadatos que respondan a las necesidades, a menudo imprevistas, del futuro.

No cabe duda de que muchos tipos de metadatos serán de una importancia crítica para el uso continuo y el acceso físico e intelectual a los recursos de información digital y a su contenido. En este sentido, los metadatos representan la piedra Rosetta que nos permitirá descodificar los objetos informáticos y transformarlos en fuentes de información dentro de los sistemas digitales del siglo XXI.

Los metadatos y la World Wide Web

Tony Gill

*Art, Design, Architecture and Media Information Gateway
and Visual Data Service, Surrey Institute of Art and Design*

Introducción

El término *metadatos* parece indicar algo extremadamente complicado y, sin embargo, el concepto subyacente a esta palabra es muy simple, al menos en el contexto de este artículo. Los metadatos son, sencillamente, datos que sirven para describir grupos de datos a los que podríamos llamar “objetos informáticos”.

Esto significa que la relación entre el objeto informático y los datos que lo describen es funcionalmente idéntica, por ejemplo, a la relación entre un libro y su registro en el catálogo de una biblioteca o entre un objeto en un museo y su registro en el sistema de organización de una colección. Por lo tanto, a lo largo de esta discusión, asumiremos que los metadatos son descripciones estructuradas de un objeto informático.

Siempre que se organizan grandes colecciones de objetos se recurre a catálogos compuestos de descripciones estructuradas. El inventario informatizado de un almacén para hacer el balance de sus existencias contiene fundamentalmente el mismo tipo de información que una base de datos para organizar las colecciones de un museo o los fondos de una biblioteca o, incluso, que las fichas de un individuo para ordenar su colección de discos compactos. Cada uno de estos registros descriptivos es un intento de captar de manera concisa y razonable las características distintivas, o esenciales, del objeto que describen.

En principio, la estructura de los registros descriptivos de una base de datos está diseñada para simplificar la búsqueda y para clasificar y visualizar la información de modo que sea más fácil usar y gestionar efectivamente una colección. El diseño de esta estructura depende de como se gestione la colección, del grado de homogeneidad entre los objetos de la colección y de la capacidad del diseñador de la base de datos a la hora de captar y estructurar la esencia de los objetos.

El crecimiento constante de la World Wide Web

No hay nadie que tenga una memoria sin límites y, por lo tanto, la importancia de los catálogos descriptivos para gestionar y acceder a una colección aumenta en proporción directa al tamaño de la misma. La falta de un catálogo de este tipo es una de las mayores desventajas de la World Wide Web, la vasta red de páginas hipermedia que sin duda constituye la colección de objetos informáticos más grande y de crecimiento más rápido del mundo.

Es difícil establecer con exactitud el índice de crecimiento de la Web, pero incluso calculando por lo bajo se obtienen unas cifras alarmantes; el Web Growth Summary⁵ de Matthew Gray sugiere que el número de sitios de la Web en el Internet ha pasado de 130 en junio de 1993 a unos 650.000, aproximadamente, en enero de 1997, lo que representa un incremento del 500.000 por ciento en 42 meses. Aun así, este cálculo está basado en una definición muy restringida de lo que es un sitio de la Web, según la cual se trata de una colección de documentos provistos de URL (Uniform Resource Locators) que empiezan con un hostname único:

Es decir, <http://www.mit.edu/people/mkgray/> y <http://www.mit.edu/madlibs> forman parte del mismo sitio, pero un documento <http://Web.mit.edu/> es un sitio separado.⁶

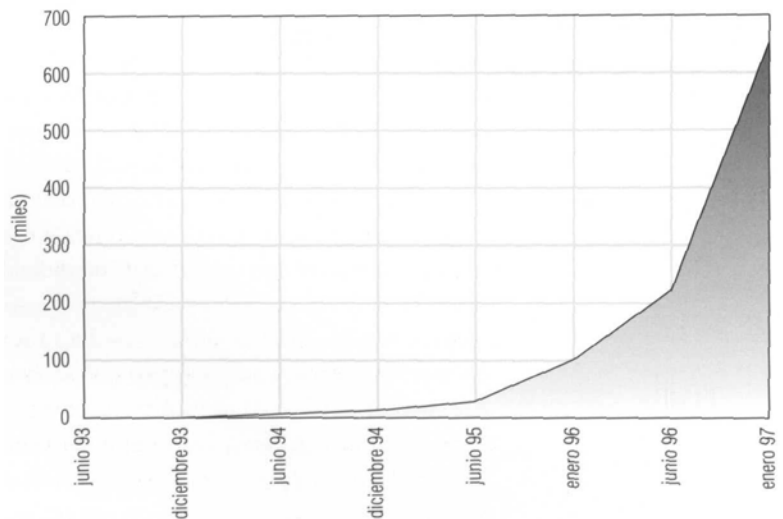


Figura 2 Crecimiento del número de sitios en la World Wide Web

⁵ Datos facilitados por Matthew Gray, Massachusetts Institute of Technology (<http://www.mit.edu/people/mkgray/net/>).

⁶ Ibid.

Esto significa que la figura de 650.000 es un cálculo muy por lo bajo del crecimiento real de la Web; los Internet Service Providers, por ejemplo, pueden proporcionar espacio en la Web a cientos, o quizá miles, de clientes en un único Web host con el mismo hostname para todos y, sin embargo, probablemente se considerarían sitios individuales.

Buscar agujas en un pajar global

Lamentablemente, ni el Internet ni la World Wide Web fueron originalmente diseñados con la idea de catalogar su contenido; el conjunto de protocolos para redes interconectadas TCP/IP, que posibilita el funcionamiento de la infraestructura básica del Internet, es solamente un vehículo para transportar paquetes de datos de un punto a otro de la forma más rápida y fiable, mientras que el HyperText Transfer Protocol (HTTP) se encarga de la entrega de información hiperenlazada la World Wide Web.

Esto significa que los protocolos existentes no proporcionan ningún soporte dedicado a la localización de los recursos de información disponibles en la red. Esta situación tan deplorable es semejante a la visión de Memex, una forma de conocimiento democrático global propuesto por Vannevar Bush en 1945. La desilusión de los expertos en hipertexto con la World Wide Web queda reflejada en un comentario de Ted Nelson (quien acuñó el término *hypertext*, por primera vez, en 1965), durante la conferencia Hypertext 97, citado a continuación:

La reacción de la comunidad de investigadores de hipertexto respecto a la World Wide Web es semejante a la de quien de pronto se da cuenta de que tiene un hijo ya mayor. Y es un delincuente.⁷

No es sorprendente que los primeros instrumentos diseñados para la localización de recursos en el Internet y para el establecimiento de un cierto orden en su inmenso caudal de información aparecieran tras la introducción de los primeros navegadores a comienzos de 1990; por ejemplo, Tim Berners-Lee, poco después de haber inventado la Web, fundó la WWW Virtual Library, mientras que Yahoo!, Lycos y Webcrawler aparecieron en 1994.

Los instrumentos de que actualmente dispone el usuario para encontrar recursos en la Web son mucho más poderosos que sus predecesoras de 1994—tienen necesariamente que serlo dado el crecimiento

⁷ Ted Nelson, durante la celebración de Hypertext 97, la Eighth ACM Conference on Hypertext, Southampton, del 6 al 11 de abril de 1997. Fuente: Trip Report on UK Web Focus Conference Center de Nick Gibbin en <http://www.ukoln.ac.uk:8002/conference/>

explosivo tanto de la información disponible como del número de usuarios que acceden a ella. Sin embargo, aún ahora, sólo hay dos tipos principales de instrumentos para la localización de recursos en la Web: los directorios y los mecanismos de búsqueda.

Los directorios son listas de recursos de la Web creados por individuos que seleccionan, catalogan y clasifican los recursos que consideran más útiles para sus clientes. Los directorios pueden ser de tipo general, como por ejemplo la World Wide Web Virtual Library⁸ y Yahoo!⁹ o pueden estar especializados en temas particulares, como por ejemplo el Art, Design, Architecture & Media Information Gateway (ADAM)¹⁰ y la Edinburgh Engineering Virtual Library (EEVL).¹¹ Generalmente, los directorios permiten el acceso a sus enlaces tanto buscando como navegando por un conjunto jerárquico de encabezamientos temáticos.

Los mecanismos de búsqueda, a menudo llamados “spiders”, “crawlers” o “robots”, son sistemas automatizados que cruzan la Web, visitando sitios continuamente, guardando copias de los recursos, así como de sus localizaciones, con el objetivo de acumular un inmenso catálogo de páginas completamente clasificadas. Por lo general, son una poderosa ayuda para la búsqueda y producen resultados enormes, los cuales se clasifican por orden de importancia para que puedan ser realmente útiles.

Sin embargo, tanto los directorios como los buscadores presentan graves problemas; debido al papel del individuo en el proceso de clasificación e indexación, los directorios son muy eficientes en la búsqueda de temas generales y, frecuentemente, proporcionan enlaces a información de alta calidad. No obstante, es un trabajo intensivo y costoso que no puede avanzar al ritmo que requeriría una puesta al día constante de toda la Web—dejando al margen el hecho de que la mayor parte de la información contenida en la Web es de carácter efímero y, por lo tanto, sería un esfuerzo inútil catalogarla.

Otro problema relacionado con el método de catalogación manual de la Web surge en el momento de decidir con qué nivel de especificidad hay que describir los recursos; ¿deberían crearse descripciones globales de todo un sitio de la Web o sería mejor catalogar cada página individualmente? Evidentemente, siempre habrá que tener en cuenta los costes y los beneficios de una y otra opción.

Por otra parte, los mecanismos de búsqueda no pueden proporcionar un índice amplio y actualizado del contenido de la Web ni tampoco

⁸ WWW Virtual Library: <http://vlib.stanford.edu/Overview.html>

⁹ Yahoo!: <http://www.yahoo.com/>

¹⁰ ADAM: <http://adam.ac.uk/>

¹¹ EEVL: <http://www.eevl.ac.uk/>

pueden garantizar que el usuario encuentre la información que busca, aunque ésta haya sido previamente catalogada.

- Los crawlers de la Web están completamente automatizados, lo que significa que los recursos de la Web se seleccionan por medio de software, sin intervención humana, y son, por lo tanto, de calidad variable.
- Buscar en bancos de datos enormes e indexados automáticamente produce a menudo resultados tan cuantiosos que, frecuentemente, son inútiles a pesar de que los instrumentos para la extracción de información, los procedimientos para su clasificación por orden de importancia y los algoritmos de inteligencia artificial sensibles al contexto sean cada vez más sofisticados.
- La información en la Web se genera cada vez más de forma improvisada (“*on the fly*”) desde bancos de datos de referencia (a menudo denominados “la Web sumergida”), que están más allá del ámbito indexado en el que operan los crawlers.
- A medida que aumente el volumen de información en la Web, la red requerirá un ancho de banda (la capacidad de llevar información) superior a los niveles máximos aceptables; ya hace algunos años que existe una ética, o “código de conducta”, para la Web.

La catalogación de la Web

Aunque a primera vista parezca que los directorios y los mecanismos de búsqueda tengan problemas distintos, un análisis más detallado demuestra que la mayoría de estos problemas son el resultado de ambiciones insostenibles a largo plazo; la Web ha crecido ya demasiado como para que ninguna organización o servicio en particular pueda catalogarla, tanto si se generaran sus índices por medio de ordenadores como si se hiciera manualmente.

Cualquier solución al problema de cómo descubrir los recursos de la Web tiene que basarse en un tipo de catálogo *distribuido* (irónicamente, la WWW Virtual Library usa precisamente un modelo distribuido; sin embargo, los esfuerzos altruistas de su personal voluntario no bastan para seguir el ritmo de crecimiento de la Web).

Los protocolos técnicos necesarios para crear sistemas distribuidos, tales como Z39.50 y WHOIS++, ya existen—la interoperabilidad a nivel técnico ya no es un problema. Lo que se necesita urgentemente son estándares más abstractos para la estructura y el contenido de la información que permitan la interoperabilidad a nivel semántico. Sin embargo, antes de determinar la estructura y el contenido más efectivo para los

metadatos de la Web, hay que establecer primero las operaciones que tendrán que realizar.

La aplicación más corriente para los metadatos de la Web se llama generalmente “descubrimiento de recursos”, y recibe este nombre porque en ella los metadatos sirven para ayudar a los usuarios a encontrar la información que buscan; la existencia de descripciones consistentes, exactas y bien estructuradas de los recursos de la Web permitiría mayor precisión en la busca y una clasificación más rigurosa de los resultados obtenidos según su relevancia.

Los metadatos también podrían servir para dar breves descripciones o evaluaciones de los recursos hallados, lo cual ayudaría a los usuarios a determinar su utilidad antes de examinarlos individualmente, o incluso podrían indicar la existencia de restricciones de acceso o de derechos sobre el uso de la información. Habría que discutir, por supuesto, la cuestión de si estas aplicaciones deberían ser una parte intrínseca del proceso de descubrimiento de recursos o bien si deberían ser aplicaciones separadas de los metadatos de la Web.

Los metadatos se usan también en la gestión y la administración de recursos digitales; este tipo de “metadatos administrativos” es esencial para mantener al día los recursos de la Web, por ejemplo, o para asegurar que esos recursos no estén sujetos a restricciones que prohíban su distribución a través del Internet.

Evidentemente, tanto el contenido como la estructura de cualquier registro que describa un recurso de la Web tienen que captar la esencia del mismo y, a la vez, tienen que facilitar las varias operaciones para las que fue diseñado. Lamentablemente, es aquí donde las complejidades de la vida real penetran el mundo de la técnica; con una colección enorme de objetos, capaz de ampliar y profundizar el conocimiento humano, y con decenas de millones de usuarios, el número de aplicaciones posibles para los metadatos de la Web es prácticamente ilimitado. Por consiguiente, a pesar de los esfuerzos internacionales realizados hasta el momento, no se ha conseguido llegar a un consenso respecto al contenido y la estructura más adecuada para los metadatos de la Web. Algunos de los estándares descriptivos más importantes que han surgido como resultado de las investigaciones sobre los metadatos se describen en este artículo y en otras secciones de la presente publicación.

Una de las consecuencias más interesantes de las investigaciones sobre metadatos llevadas a cabo internacionalmente es que la catalogación—considerada hasta ahora como un arte arcano practicado solamente por bibliotecarios, curadores de museos y archiveros—se está convirtiendo en una cuestión que atañe a una comunidad cada vez más amplia. Si bien hay, sin duda, mucho que aprender de los métodos tradicionales de catalogación, el carácter multidisciplinario, global, multilingüe

y multicultural de la Web plantea problemas que exigen nuevos métodos y distintas soluciones.

Se acepta cada vez más, por ejemplo, la importancia de los vocabularios controlados y de los sistemas formales de clasificación para catalogar los recursos de la Web—algo que los catalogadores profesionales han tenido en cuenta durante años (a pesar de que la misma diversidad de información de la Web ponga en cuestión la capacidad de las taxonomías existentes para organizar la suma del conocimiento humano).

Sin embargo, la magnitud de la Web como espacio de información exigirá nuevas aplicaciones a los procedimientos habituales, tales como la existencia de tesauros en programas de software para ampliar automáticamente las opciones del usuario y poder así incluir sinónimos o incluso traducciones de los términos a lenguas alternativas, o permitir correspondencias entre diferentes terminologías y sistemas de clasificación. Asimismo, el hecho de que diversos tipos de vocabularios y de sistemas de clasificación tendrán que coexistir en el mismo espacio significa que los ordenadores deberán identificar las fuentes de autoridad para términos o marcas de clase (*class marks*); en consecuencia, serán necesarios registros que aseguren que las etiquetas empleadas para la identificación de las varias autoridades sean únicas y sin ambigüedades.

El futuro catalogador de la Web tendrá que decidir también el nivel de detalle más apropiado para la descripción de los recursos: ¿Cómo debería ser de detallado el registro de un recurso de la Web? ¿Cuántos registros de catalogación deberían crearse para cada recurso? Las crecientes expectativas de los usuarios respecto a la accesibilidad de los datos, junto con la flexibilidad y la diversidad de la información en hipertexto, hacen que la comparación entre la catalogación de recursos de la Web y la catalogación de recursos bibliográficos sea sólo parcialmente válida. Poco satisfecho con la fórmula tradicional, “autor, título, palabra clave”, que ofrecen los catálogos de las bibliotecas, el usuario espera poder buscar por palabra clave tanto dentro de las subsecciones de los títulos como en el mismo texto de los recursos de la Web.

Con el objetivo de superar estos problemas, el Art, Design, Architecture & Media Information Gateway está colaborando con un productor de software para combinar el método de directorio con los mecanismos de búsqueda; un grupo de bibliotecarios especializados está creando descripciones de alta calidad a nivel de sitio de recursos cuidadosamente seleccionados, que se incrementarán después con una indexación de todo el texto del sitio por medio de un crawler. Otra dificultad conceptual importante surge de la necesidad de describir las relaciones de los recursos interconectados con otros objetos informáticos: ¿Qué es lo que deberían describir los metadatos exactamente? En sentido estricto, los metadatos deberían describir las propiedades de un objeto que es en sí mismo un

conjunto de datos como, por ejemplo, una página Web, una imagen digital o una base de datos—esto es análogo a lo que hace un bibliotecario cuando cataloga un objeto tangible. Para descubrir recursos interconectados, sin embargo, estas propiedades no son particularmente útiles; por ejemplo, un investigador interesado en encontrar imágenes de obras de arte famosas en la Web, generalmente buscaría usando las propiedades de la obra de arte original (p. ej., CREADOR = Picasso, FECHA = 1937), no las propiedades de las copias digitales de la obra (p. ej., CREADOR = Scan-U-Like Imaging Labs Inc., FECHA = 1998).

Tanto el problema de nivel de detalle (*granularity*) como el de las copias u objetos digitales (*surrogacy*) surgen de la necesidad de describir relaciones entre objetos distintos (que no van a existir todos en la Web) en los varios registros en los que se describen estos objetos; por ejemplo, un registro que describa una página Web debería indicar a qué sitio pertenece y, asimismo, la imagen de una pintura de Picasso en la Web debería identificar la pintura de la que procede.

Por supuesto que ninguno de los problemas descritos hasta ahora es una novedad—durante décadas, tanto bibliotecarios como curadores y archiveros se han enfrentado a la tarea, aparentemente imposible, de “recrear el mundo” para poder describir los recursos de información. No obstante, la urgente necesidad de catalogar la Web ha hecho que estas cuestiones, fundamentalmente epistemológicas, se conviertan en el foco de interés de una comunidad nueva y mucho más amplia.

Estándares para metadatos en la Web

Para que los metadatos sean útiles en la construcción de un catálogo distribuido de la Web, es esencial que tanto la estructura como la sintaxis de la información se ajusten a estándares de amplia aceptación, de modo que la semántica de los registros del catálogo pueda ser correctamente interpretada por los distintos ordenadores que integran la red del catálogo distribuido.

La presente publicación contiene una sección en la que se ofrece una perspectiva general de los varios estándares para metadatos usados habitualmente. Aquí sólo vamos a discutir tres grupos de estándares para metadatos, estrechamente relacionados entre sí, que son de particular importancia en el ámbito de la Web: las etiquetas <META> de AltaVista, el Dublin Core y el Resource Description Framework.

Etiquetas <META> de AltaVista

El popular mecanismo de búsqueda AltaVista sirve de apoyo a dos elementos muy simples, *Palabras clave* y *Descripción*, que pueden insertarse en

recursos de la Web usando la etiqueta HTML<META>. La información suministrada por las *Palabras clave* contribuye a la clasificación de la información según su relevancia, mientras que la información que proporciona la *Descripción* se usa al visualizar los resultados para obtener un resumen más preciso de un recurso en particular.

Dublin Core

El Dublin Core Metadata Element Set (conocido también como “Dublin Core” o simplemente “DC”) es un conjunto de quince elementos de información que se pueden usar para describir una amplia gama de recursos en el Internet, lo que permite una forma simple de descubrimiento interdisciplinario de recursos. Estos quince elementos son: *Contributor* (Contribuidor), *Coverage* (Ámbito), *Creator* (Creador), *Date* (Fecha), *Description* (Descripción), *Format* (Formato), *Identifier* (Identificante), *Language* (Lengua), *Publisher* (Editor), *Relations* (Relaciones), *Rights* (Derechos), *Source* (Origen), *Subject* (Tema), *Title* (Título) y *Type* (Tipo).

Tanto los elementos del DC como su significado fueron creados y modificados por un grupo de bibliotecarios, especialistas en información y especialistas temáticos, a través de un proceso en el que cada decisión se sometió a consenso y del que formaron parte seis talleres internacionales.¹²

Desde el primer momento, el desarrollo del Dublin Core vino determinado por las siguientes directrices:

- Los elementos tienen que ser fáciles de entender y de usar, de modo que cualquier individuo que haya creado recursos interconectados debería poder describirlos sin necesidad de un período extensivo de capacitación.
- Cada elemento es opcional y repetible.
- Los elementos deberían ser internacionales e interdisciplinarios en ámbito y en aplicabilidad.
- El conjunto de elementos debería ser ampliable, para permitir modificaciones relacionadas con una disciplina particular o con funciones específicas.
- La aplicación estratégica más importante del conjunto sería para las descripciones insertadas de recursos de la Web creadas por los autores del recurso, lo que implica una sintaxis que debería poder adaptarse a las etiquetas HTML <META>.

¹² Véase <http://purl.org/DC/about/workshop.htm>.

Los primeros usuarios del Dublin Core se encontraron con problemas parecidos a los discutidos en la sección anterior y, como consecuencia, se introdujeron una serie de modificaciones al conjunto inicial de elementos:

- El *Warwick Framework* es una arquitectura de envase conceptual para paquetes heterogéneos de metadatos; las aplicaciones SGML y MIME del Warwick Framework ya se han puesto en práctica, pero quizá la contribución más importante haya sido la formalización de los requisitos necesarios para la creación del Resource Description Network.
- Los *Canberra Qualifiers* son un conjunto de tres calificadores (SUBELEMENT, SCHEME y LANG) que se pueden usar para perfeccionar la semántica del conjunto de elementos DC y obtener así información más precisa.
- Un acuerdo según el cual la única solución práctica a los problemas relacionados con el nivel de detalle y las copias digitales es usar “conjuntos” o “paquetes” separados de metadatos para cada objeto particular (ya sea una pieza o una colección, una red de recursos u otra cosa) y describir las relaciones entre ellos por medio de una lista enumerada de tipos de relaciones.

Sin embargo, el resultado más significativo del Dublin Core es, indudablemente, el consenso internacional e interdisciplinario al que se está llegando.

Resource Description Framework

El Resource Description Framework (RDF), asociado a las actividades sobre metadatos del World Wide Web Consortium, es una aplicación de metadatos del XML (Extensible Markup Language), considerado por la mayoría de expertos el sucesor del HTML y el futuro lenguaje de la Web. Su evolución se vio influida por proyectos anteriores como PICS (Platform for Internet Content Selection), por la iniciativa Dublin Core/Warwick Framework y por el trabajo sobre metadatos llevado a cabo por los productores más importantes de software, como Microsoft y Netscape.

El RDF facilitará una arquitectura flexible para la utilización de paquetes de metadatos diversos, creados para una función específica, que pueden ser procesados por máquinas. También admitirá la integración de “firmas” digitales, gracias a las cuales las organizaciones podrán sancionar, o “firmar”, un paquete de metadatos como garantía de que éstos representan una descripción honesta y auténtica de un determinado recurso de la red.

Metamenazas y metaoportunidades

Uno de los problemas más importantes con la arquitectura HTML actual para insertar metadatos en los recursos de la Web es que es susceptible de abuso. A medida que ciertas organizaciones comerciales de dudosa reputación han visto las ventajas de aparecer en la parte superior de una lista de resultados en la Web, se han incrementado los intentos de manipular el algoritmo de acceso a los datos, por ejemplo, repitiendo palabras clave cientos de veces en los metadatos o usando palabras clave explícitamente sexuales en los mismos. Esta técnica, conocida bajo el nombre de *spamming*, está creando una falta de incentivos para los autores de la Web que desean insertar metadatos descriptivos—en lugar de apoyar los metadatos insertados, los mecanismos de búsqueda tienen que adoptar una actitud cada vez más escéptica respecto a ellos a fin de frustrar los intentos de los *spammers*.

Sin embargo, la posibilidad de una autenticación digital creará nuevas oportunidades para proveedores de información fiables, tales como bibliotecas, museos y agencias oficiales de información, que recuperarán así su papel de guardianes de recursos de alta calidad en la World Wide Web del siglo XXI.

Correspondencias entre metadatos e interoperabilidad: ¿Qué significa todo esto?

*Willy Cromwell-Kessler
The Research Libraries Group*

Una de las mayores ventajas de las nuevas redes informáticas interconectadas es la posibilidad de integrar recursos de información distintos pero complementarios. Estos recursos pueden abarcar desde la información descriptiva en los bancos de datos de citas y catálogos hasta colecciones de recursos de información íntegros tales como imágenes o textos completos digitalizados. Generalmente, estos recursos se exploran por medio de datos descriptivos, específicamente estructurados—los “metadatos”. Los bancos de datos de citas y catálogos constituyen tipos específicos de metadatos, mientras que los objetos informáticos íntegros, es decir, aquellos que contienen la información completa, pueden ir acompañados por ellos. Este tipo de metadatos refleja generalmente la naturaleza de los recursos descritos, así como los objetivos y los objetos propios de la comunidad a la que sirve y, en consecuencia, puede que esos metadatos difieran significativamente de un recurso a otro. Una manera de navegar entre distintos recursos es a través de correspondencias, obtenidas por medio del análisis de los sistemas de metadatos y del establecimiento de asociaciones entre ellos. La creación de correspondencias entre los distintos sistemas de metadatos es el modo más eficaz de reconciliar sus diferencias y de facilitar la interoperabilidad entre sistemas para llegar así al acceso integrado.

Los sistemas de metadatos pueden diferenciarse entre sí tanto por el contenido como por la estructura. El contenido puede ser distinto en términos de las reglas que rigen su formulación o en cuanto al lenguaje, ya que el mundo electrónico es cada vez más internacional. Sin embargo, el mayor obstáculo a la integración es la estructura de los distintos sistemas de metadatos, que pueden estar constituidos por diversos elementos funcionando a diferentes niveles y designados de manera distinta. A menudo se encuentran sistemas divergentes que compiten dentro de un mismo terreno, debido al desarrollo aislado de sistemas de metadatos y al hecho de que la uniformidad no se consideraba necesaria cuando no se tenía acceso a las redes de información.

En algunos casos quizá sea posible reconciliar estos sistemas de metadatos redundantes y desarrollar estándares diseñados para una disciplina específica, o bien estándares interdisciplinarios. Sin embargo, esta solución puede que sea sólo hipotética, ya que la conversión retrospectiva de datos preexistentes es lenta y costosa. Si no existe un estándar único, la integración supone la “traducción” de un sistema de datos estructurados a otro. Desde este punto de vista, las correspondencias entre los sistemas de metadatos funcionan como una especie de piedra Rosetta, puesto que son la clave de la traducción automática de un sistema a otro. Un ejemplo común de estas correspondencias son las implementaciones Z39.50, que requieren la conversión de datos antes de permitir que el sistema empleado por un usuario pueda traducir y extraer datos de manera conforme a sus propios protocolos.

Como sabe todo lingüista, la traducción de un texto de una lengua a otra nunca es exacta en el sentido estricto de la palabra. Asimismo, debido a que los sistemas de metadatos son distintos en cuanto a la amplitud, la profundidad, el énfasis y el ámbito que cubren, las correspondencias entre ellos son, a menudo, aproximadas. Algunas de las cuestiones más difíciles de resolver son las siguientes:

- Dos o más conceptos en un sistema pueden estar representados por un elemento único en otro sistema. Esto no representa un problema cuando vamos de un sistema más inclusivo a otro menos inclusivo, pero el movimiento inverso puede ser más difícil o incluso imposible.
- Puede que ciertos elementos específicamente delineados en un sistema de metadatos se puedan proyectar a otro solamente en el campo de “notas” o “comentarios”. Esto contribuye también a que el movimiento inverso sea problemático.
- Puede que no haya ningún equivalente en un sistema para conceptos expresados en otro.
- Puede haber una tensión de un sistema a otro entre los datos descriptivos encargados de la visualización y los datos que sirven para la indexación y/o el acceso a la información.

Tal como sugieren los problemas mencionados arriba, el proceso de correspondencia de un sistema a otro puede plantear problemas muy complejos. Puesto que el objetivo de las correspondencias es facilitar el descubrimiento y el acceso a los recursos a través de una amplia gama de fuentes de información, es mejor que estas decisiones se tomen de manera uniforme y que no difieran de una a otra organización. Consecuentemente, es importante que se desarrollen y se mantengan registros de metadatos acreditados y exhaustivos que ofrezcan definiciones del signifi-

cado, de la estructura y del contenido de diversos sistemas y que, asimismo, incluyan correspondencias formales entre los elementos comunes cuando se considere adecuado.

Sólo pueden mantener auténticos registros de metadatos las organizaciones que se ocupan de formular estándares formales como la International Organization for Standardization (ISO), que ha empezado a estudiar la cuestión de la especificación y la estandarización de datos (véase ISO 11179). Sin embargo, en espera de que estos registros formales y exhaustivos puedan empezar a desarrollarse, algunas organizaciones han iniciado la creación y el mantenimiento de correspondencias menos formales que, de hecho, funcionan como registros. Dos ejemplos de estos sitios informales son los mantenidos por la UK Office for Library and Information Networking (UKOLN) (<http://ukoln.bath.ac.uk/metadata/interoperability>) y la Library of Congress (<http://www.loc.gov/marc>).

Correspondencias RLG/Getty de metadatos

La correspondencia incluida en la presente publicación es una versión preliminar de la que el Research Libraries Group está desarrollando en colaboración con el Getty. Esta correspondencia estará a disposición del público en la Web (<http://www.rlg.org>) y será actualizada y ampliada periódicamente; se centrará en sistemas de metadatos para información sobre el patrimonio cultural. Los sistemas de metadatos representados actualmente en esta correspondencia están especializados en la descripción de obras de arte, objetos de museos y otra información relativa al patrimonio cultural, así como en algunos conjuntos de elementos generales:

- Las *Categorías para la descripción de obras de arte* ofrecen una taxonomía de información sobre el patrimonio cultural muy articulada e inclusiva. Este estándar, que ha sido calificado como “la madre de los estándares para metadatos sobre el patrimonio cultural”, proporciona directrices muy detalladas para la descripción de los objetos artísticos y de sus copias digitales.
- El Object ID consiste en una lista de términos que definen la información mínima esencial para poder seguir la pista de objetos artísticos perdidos o robados.
- El CIMI Schema, derivado del CIDOC Data Model y de las *Categorías para la descripción de obras de arte*, es de capital importancia, ya que define una lista de elementos que probablemente existen en la mayoría de bancos de datos sobre el patrimonio cultural, y ha inspirado la creación de instrumentos como el Z39.50 para el acceso a dicha información.

- La *Guide to the Description of Architectural Drawings* de la FDA/ADAG (Foundation for Documents of Architecture/Architectural Drawing Advisory Group) proporciona directrices para la descripción y el acceso a información sobre documentos de arquitectura.
- Las categorías VRA Core, que también están basadas en las *Categorías para la descripción de obras de arte*, son metadatos para la descripción de obras de arte y de sus copias digitales. Este proyecto, que aún está en fase de desarrollo, fue iniciado por el Data Standards Committee de la Visual Resources Association y se propone establecer los elementos necesarios, así como el método adecuado, para describir la documentación visual del patrimonio cultural y de los objetos de museos. El proyecto VISION (Visual Resources Sharing Information Online) ha sido diseñado para experimentar la aplicación del VRA Core en la Web.
- El REACH (Record Export for Art and Cultural Heritage) proporciona un formato de ficha común para la exportación de información sobre la gestión de colecciones; el objetivo es crear un banco de pruebas para registros de objetos de museo. Este estándar fue diseñado para facilitar la localización y el uso repetido de la información y, asimismo, para potenciar el desarrollo de estándares que faciliten el intercambio de información entre los museos.
- El USMARC no es estrictamente un sistema diseñado para la información sobre el patrimonio cultural. Sin embargo, se incluye aquí porque tiene una amplia aceptación en los Estados Unidos y se encuentra con frecuencia en el repertorio del patrimonio cultural. El formato MARC (MACHINE-Readable Cataloging) es una estructura de datos para codificar registros de libros, material de archivo, manuscritos, material visual, etc. que se ha usado en bibliotecas y archivos durante muchos años.
- El Dublin Core, que define metadatos para encontrar recursos en la World Wide Web, tiene también una aplicación más general. Se incluye aquí porque existe un gran interés en explorar el papel que podría tener en la información cultural contenida en la Web.

Estándares para metadatos: Comparación

En las tablas que siguen a continuación se comparan varios estándares para metadatos, subrayando sus similitudes y sus diferencias en cuanto al ámbito de aplicación. Cada uno de estos estándares representa, por así decirlo, un “punto de vista”—mientras que las *Categorías para la descripción de obras de arte* proporcionan directrices amplias e inclusivas para la descripción de objetos artísticos desde una perspectiva académica o de investigación, el Object ID codifica un conjunto mínimo de datos esenciales (diez en total) para la protección de un objeto contra el robo o el tráfico ilícito. El CIMI Schema define datos para que los museos puedan elaborar información muy detallada. Las directrices FDA se centran en documentos de arquitectura, mientras que el MESL y las VRA Core Categories proporcionan datos tanto para el objeto como para copias del mismo. El objetivo de los datos facilitados por el proyecto REACH consiste en proporcionar una estructura para exportar datos digitales procedentes de distintos sistemas y, asimismo, analizar el valor que tiene esta estructura para la investigación de la información unificada resultante. El USMARC es un estándar para datos usado durante mucho tiempo en las bibliotecas, mientras que el Dublin Core facilita elementos básicos de información, o categorías de metadatos, para mejorar la indexación y la búsqueda de los recursos de la World Wide Web.

Algunos de los estándares para metadatos del patrimonio cultural que no se han incluido en la tabla son los siguientes: SPECTRUM, un estándar creado por el Museum Documentation Association (MDA) que define veinte procedimientos llevados a cabo en museos, así como el tipo de datos que requieren; el CIDOC Data Model, desarrollado por el International Council of Museums Documentation Committee; el CIDOC Guidelines for Museum Object Standards; AFRICOM, un estándar del International Council of Museums; el diccionario de datos del AMICO (Art Museum Image Consortium). Todos estos estándares tienen su correspondencia en las directrices trazadas en *Categorías para la descripción de obras de arte*.

Correspondencias entre estándares para metadatos

CDWA	Object ID	CIMI Schema	FDA	MESL
OBJECT/WORK (core)			Document Classification-Catalogue Level (core) Document Classification-Group Type	
Object/Work-Type (core)	Type of Object	objectName	Document Classification-Document Type (core) Purpose-Purpose (Broad) (core) Purpose-Purpose (Narrow)	Object Type/Object Class/Object Name (required) Concepts/Functions
Object/Work-Components		quantity	Document Classification-Extent	Parts/Pieces
CLASSIFICATION (core)				
ORIENTATION/ARRANGEMENT				
TITLES OR NAMES (core)	Title	objectTitle bibliographicTitle	Group/Item Identification-Repository Title Group/Item Identification-Descriptive Title (core) Group/Item Identification-Inscribed Title	Object Title/Caption (required)
STATE				Edition/State
EDITION				Edition/State
MEASUREMENTS				
Measurements-Dimensions	Measurements	dimensions	Physical Characteristics-Dimensions Description	Dimension/Extent-Quantity-Unit
Measurements-Dimensions-Type			Physical Characteristics-Height	
Measurements-Dimensions-Value			Physical Characteristics-Width Physical Characteristics-Depth	
Measurements-Dimensions-Unit			Physical Characteristics-Unit of Measurement	
Measurements-Scale			Physical Characteristics-Scale Description Physical Characteristics-Scale	
MATERIALS AND TECHNIQUES				
Materials and Techniques-Description	Materials and Techniques		Physical Characteristics-Technique, Medium, and Support Description	
Materials and Techniques-Processes or Techniques		processTechnique		Creation Technique/Method/Process

VRA Core Categories	REACH	USMARC	Dublin Core
W1. Work Type	Field #1: Type of Object	655 Genre-Form	Type [or, Source.Type]
		300a Physical Description-Extent	Subject [or, Source.Subject]
			Description [or, Source.Description]
W2. Title	Field #4: Object Name/Title	24Xa Title and Title-Related Information	Title [or, Source.Title]
		562c Copy and Version Identification Note-Version Identification	Description [or, Source.Description]
		250 Edition Statement	
W3. Measurements	Field #7: Dimensions	340b Physical Medium-Dimensions 300c Physical Description-Dimensions	
W5. Technique	Field #5: Techniques/Processes	340d Physical Medium-Information Recording Technique	

CDWA	Object ID	CIMI Schema	FDA	MESL
Materials and Techniques-Processes or Techniques-Name			Physical Characteristics-Technique	
Materials and Techniques-Processes or Techniques-Implement			Physical Characteristics-Technique	
Materials and Techniques-Materials [Materials and Techniques-Materials-Role and Materials and Techniques-Materials-Name]		materialMedium	Physical Characteristics-Medium Physical Characteristics-Support	Materials/Medium Support
FACTURE				
PHYSICAL DESCRIPTION	Distinguishing Features	physicalDescription		
INSCRIPTIONS/MARKS	Inscriptions and Markings	inscriptionMark	Physical Characteristics-Inscription Description Physical Characteristics-Inscription	Marks/Inscriptions
CONDITION/EXAMINATION HISTORY	Distinguishing Features	condition		
CONSERVATION/TREATMENT HISTORY	Distinguishing Features			
CREATION (core)				
Creation-Creator (core)	Maker	creatorGeneral	Origin/Maker-Responsibility Description (core) Related People/Corporate Bodies-Description	
Creation-Creator-Identity-Names (core)		creatorName	Origin/Maker-Name (core) Related People/Corporate Bodies-Name	Creator/Maker-Name (required)
Creation-Creator-Identity-Dates				
Creation-Creator-Identity-Dates/Location-Birth (core)		creatorDateofBirth		
Creation-Creator-Identity-Dates/Location-Death (core)		creatorDateofDeath		
Creation-Creator-Identity-Nationality/Culture/Race (core)		creatorNationality-CultureRace		Creator/Maker-Culture/Nationality
Creation-Creator-Role (core)		creatorRole	Origin/Maker-Role (Broad) (core) Origin/Maker-Role (Narrow) (core) Related People/Corporate Bodies-Related Role	Creator/Maker-Role
Creation-Date (core)	Date or Period	dateofOrigin	Date of Execution-Descriptive Date (core) Date of Execution-Earliest Date (core) Date of Execution-Latest Date (core)	Creation Begin Date Creation End Date

VRA Core Categories	REACH	USMARC	Dublin Core
W4. Material	Field #6: Medium/Materials	340a Physical Medium-Material Base and Configuration	Format [or, Source.Format]
		562a Copy and Version Identification- Identifying Markings	
		583l Action Note-Status	
		583x or 583z Action Notes- Nonpublic or Public Note	
W6.Creator	Field #10: Creator/Maker	1XX Main Entry 7XX Added Entry	Creator [or, Source.Creator]
	Field #11: Dates of Creator/Maker	1XXd Main Entry- Associated Dates 7XXd Added Entry- Associated Dates	
W15.Nationality/Culture	Field #12: Nationality/Culture of Creator/Maker	65X Subject Index Term 545a Biographical or Historical Data	
W7.Role		1XXe Main Entry- Relator Term 7XXe Added Entry- Relator Term	
W8.Date	Field #2: Date of Creation/Date Range	260c Imprint- Date of Publication	Date [or, Source.Date]

CDWA	Object ID	CIMI Schema	FDA	MESL
Creation-Place		placeofOrigin		Creation Place
Creation-Commission-Commissioner		association-name	Related People/Corporate Bodies-Name	
OWNERSHIP/COLLECTING HISTORY				
Ownership/Collecting History-Description		owner provenance	Provenance-Provenance Description	
Ownership/Collecting History-Owner		owner provenance	Provenance-Former Owner Name	
Ownership/Collecting History-Transfer Mode				Accession Method
Ownership/Collecting History-Credit Line		creditLine		Credit Line (required)
COPYRIGHT/RESTRICTIONS				
		copyrightRestriction	Internal Documentation-Restrictions	
STYLES/PERIODS/GROUPS/MOVEMENTS				
		stylePeriod		Concepts/Style-Period
		periodName		
SUBJECT MATTER (core)				
Subject Matter-Description	Subject	contentGeneral	Method of Representation/Point of View-Method/View Description	Concept/Subject
Subject Matter-Description-Indexing Terms (core)		subject	Method of Representation/Point of View-Method/View (Broad) Method of Representation/Point of View-Method/View (Narrow)	
Subject Matter-Identification-Indexing Terms (core)		subject	Subject/Built Work Identification-Subject/Built Work Name (core)	
CONTEXT				
Context-Architectural		associationGeneral		Associated Events, People, Organizations, Places
Context-Historical/Cultural		contextHistorical		
Context-Archaeological		contextArchaeological		
EXHIBITION/LOAN HISTORY			Exhibition History	

VRA Core Categories	REACH	USMARC	Dublin Core
	Field #3: Place of Origin/Discovery (also possibly CDWA Context-Archeological)	651 Subject Term-Geographical Name 752 Hierarchical Place Name	
		7XX Added Entry 536 Funding Information Note	
		561a Ownership and Custodial History-History	
	Field #13: Current Owner Field #17: Provenance	6XX Subject Access Fields	
		541c Immediate Source of Acquisition-Method of Acquisition	
		541ac Immediate Source of Acquisition and Method of Acquisition	
		540a Terms Governing Use and Reproduction	Rights [or, Source.Rights]
W14.Style/Period/Group/Movement	Field #9: Style/Period/Group/Movement/School	65X Subject Index Term	Description [or, Source.Description]
	Field #8: Subject Matter	520 Summary, etc.	Description [or, Source.Description]
W16. Subject		65X Subject Index Term	Subject [or, Source.Subject]
W16. Subject		65X Subject Index Term	Subject [or, Source.Subject]
W13. Original Site W14. Current Site		651 Subject Access-Geographic Name 752 Hierarchical Place Name	Coverage [or, Source.Coverage] or Subject [or, Source.Subject]
	Field #3: Place of Origin/Discovery		
W13.Original Site W14. Current Site	Field #3: Place of Origin/Discovery	651 Subject Access-Geographic Name 752 Hierarchical Place Name	Coverage [or, Source.Coverage] or Subject [or, Source.Subject]
		585a Exhibitions Note	

CDWA	Object ID	CIMI Schema	FDA	MESL
RELATED WORKS		relatedObjects relatedObjects-rendition	Related Groups/Items	
Related Works-Relationship Type				
Related Works-Identification				
RELATED VISUAL DOCUMENTATION		mrObject		
Related Visual Documentation-Relationship Type		mrObject-resourceType		
Related Visual Documentation-Image Type				
Related Visual Documentation-Image Measurements		mrObject-Description		
Related Visual Documentation-Image Ownership-Owner's Name		mrObject-publisher		
Related Visual Documentation-Image Ownership-Owner's Numbers				
Related Visual Documentation-View		mrObject-coverage		
Related Visual Documentation-View-Indexing Terms		mrObject-subjectKeyword		
Related Visual Documentation-Image Source-Name				
Related Visual Documentation-Image Source-Number		Rendition-resourceIdentifier renditionIdentifier		Accompanying Image-File Name
Related Visual Documentation-Remarks				Accompanying Image-Caption (required) Accompanying Image-Capture Data
RELATED TEXTUAL REFERENCES		relatedTextualReferences	Bibliographic References-Short Citation Bibliographic References-Page Reference	
Related Textual References-Identification				Accompanying Document-File Name
Related Textual References-Type				Accompanying Document-Type
Related Textual References-Work Illustrated			Bibliographic References-Published Reproductions	
CRITICAL RESPONSES				
CATALOGING HISTORY			Internal Documentation-Note	
Cataloging History-Language				

VRA Core Categories	REACH	USMARC	Dublin Core
	Field #20: Related Objects	580 Linking Entry Complexity Note 787 Nonspecific Relationship Entry	Relation
W18. Relationship Type		787g Nonspecific Relationship Entry- Relationship Information	
W17.Related Works		787n Nonspecific Relationship Entry-Note	
	Field #19: Electronic Location & Access		
		533a Reproduction Note- Type of Reproduction	Relation
V1.Visual Document Type		533e Reproduction Note- Physical Description	
V3.Visual Document Measurements		533e Reproduction Note- Physical Description	
V5. Visual Document Owner		533c Reproduction Note- Agency Responsible for Reproduction	
V6.Visual Document Owner Number		533n Reproduction Note- Note about Reproduction	
V7.Visual Document View Description		245p Title-Name of Part or Section of Work 505 Contents 520 Summary, etc.	
V8.Visual Document Subject		65X Subject Index Term	
V9.Visual Document Source		533n Reproduction Note- Note about Reproduction	
		035 System Control Number	
		581 Publications about Described Materials	Relation
		581 Publications about Described Materials	
		510 Citation/Reference Note	
	Field #18: Language		

CDWA	Object ID	CIMI Schema	FDA	MESL
Cataloging History-Remarks			Internal Documentation-Sources	
CURRENT LOCATION (core)				
Current Location-Repository Name (core)		repositoryName	Group/Item Identification-Repository Name (core)	Holding Institution (required)
			Group/Item Identification-Administrative Unit (core)	
Current Location-Geographic Location (core)		repositoryPlace	Group/Item Identification-Repository Geographic Location (core)	
Current Location-Repository Numbers (core)			Group/Item Identification-Group/Item ID (core)	Accession Number (required)
			Group/Item Identification-Alternate ID	
DESCRIPTIVE NOTE	Description		Descriptive Note (core)	Description
			Purpose-Purpose Description	

Nota sobre las correspondencias:

Los elementos se asocian al más alto nivel de correspondencia y se considera que abarcan todos los subelementos relevantes de los conjuntos de datos correspondientes. Los elementos alternativos del Dublin Core se especifican entre corchetes cuando el elemento al que sirven de fuente se ha usado para describir un original del cual se ha obtenido algún tipo de documento visual. Los elementos que no están entre corchetes han de interpretarse como pertenecientes a la descripción del objeto original y no a la descripción de una reproducción del mismo, en cuyo caso se indica a través de un elemento de relación.

Finalmente, puesto que los conjuntos de elementos citados arriba se crearon para satisfacer distintas necesidades y para distintos grupos de usuarios, los modelos conceptuales y semánticos que les dan coherencia no son idénticos, incluso cuando los elementos están asociados. Por lo tanto, es inevitable que en algunos casos las correspondencias parezcan algo forzadas. Tenemos mucho interés en recibir sus comentarios sobre este proyecto.

VRA Core Categories	REACH	USMARC	Dublin Core
W9.Repository Name	Field #14: Current Repository Name	535a Location of Originals/Duplicates-Custodian	
W10.Repository Place	Field #15: Current Repository Place	535bc Location of Originals-Address and Country	
W11.Repository Number	Field #16: Current Object I.D. Number	035 System Control Number	
W19.Notes	Field #21: Notes	5XX General Notes	

Conclusión

Esta publicación se debe a los esfuerzos de varios individuos que, desde distintos lugares (el sur y el norte de California, Gran Bretaña), compartiendo intereses comunes y desde perspectivas distintas, trabajaron juntos hacia un mismo objetivo. A pesar de que cada colaborador haya aportado su propio punto de vista y sus ideas al tema de los metadatos, resulta grato comprobar que existan tantos puntos de contacto entre ellos: todos los colaboradores insisten en la necesidad urgente de establecer estándares basados en un consenso entre los grupos interesados, y todos afirman que sólo el uso consistente de metadatos, cuidadosamente elaborados, permitirá que la caótica masa de datos disponibles mundialmente en las redes de información se convierta verdaderamente en una “biblioteca digital”. Dos de los colaboradores, sin comunicárselo mutuamente, usan la metáfora de la piedra Rosetta, la antigua piedra egipcia que permitió descifrar la escritura jeroglífica. Ciertamente, el uso inteligente de estándares para metadatos, cuidadosamente elaborados, puede proporcionar una especie de guía que nos permita viajar por la inmensa cantidad de información que tenemos a nuestra disposición.

Otro aspecto en el que todos insisten es el de la colaboración: en la era de la información no podemos permitirnos el lujo de trabajar aislados, sin saber lo que otros están haciendo. Debemos trabajar juntos para poder navegar por el vasto universo de información en forma digital que se abre ante nosotros.

Glosario

algoritmo [algorithm]

Una fórmula para resolver un problema. Un algoritmo es un conjunto de pasos en un orden muy específico, como una fórmula matemática o las instrucciones en un programa de ordenador. Los procesos de los ordenadores están gobernados por algoritmos.

análogo de codificación [encoding analog]

Una correspondencia entre elementos de metadatos específicos en un DTD SGML (véase más abajo) y un equivalente en un conjunto de metadatos alternativos.

ancho de banda de la red [network bandwidth]

Esta expresión procede del término usado para describir la medida o el "ancho" de las frecuencias que llevan comunicaciones analógicas tales como la radio y la televisión. En el caso del Internet, el ancho de banda es esencialmente una medida de la velocidad de transferencia de datos.

atributo [attribute]

Una propiedad de un elemento de dato que ha sido nombrada, como "nivel de descripción"; los valores permitidos para un atributo pueden incluir "colección", "ficha", o "item".

base de datos de referencia [back-end database]

Una base de datos que proporciona datos para un sistema de información, y es distinta de los componentes de presentación o de interfase de ese sistema.

CGI script

Un programa de ordenador, frecuentemente escrito en C, Perl o en un "shell script",¹³ que usa el estándar interfaz de pasarela común [Common Gateway Interface] (GCI) y

¹³ Un archivo de comandos para ser ejecutados por el shell, es decir, el procesador de comandos del sistema operativo.

proporciona una interfase interactiva entre el usuario, o una aplicación de ordenador externa, y un servidor del World Wide Web. El GCI script se usa comúnmente para crear formas que permitan a los usuarios someter información a un servidor de la Web.

codificación de información [encoding ("marking up") information]

Una forma en que el creador de un objeto digital puede estructurar y codificar texto u otros datos para que puedan ser manipulados por un ordenador o por un usuario, transmitidos o buscados a través de una red o visualizados tal como los vio su realizador.

descubrimiento de recursos [resource discovery]

El proceso de buscar una información específica en el Internet.

directorio del Internet [Internet directory]

Una lista organizada temáticamente de enlaces descriptivos a sitios del Internet, generalmente creada por individuos que han clasificado los sitios según su contenido.

dominio de nombre [domain name]

La dirección que identifica un sitio del Internet o de otra red. Típicamente, el dominio de nombre consiste, como mínimo, de dos partes: el nombre de la compañía, institución u otra organización y el subdominio más alto, e.g., *www.gii.getty.edu*, donde *getty* es la compañía u organización y *gii* es el subdominio más alto. En otras palabras, el dominio de nombre es un espacio para nombres jerárquicamente estructurado que funciona como un alias para cierto ámbito de direcciones del Internet (IP). Por ejemplo, el dominio "ac.uk" identifica una serie de direcciones IP dentro del dominio del Internet académico en el Reino Unido.

DTD

Document Type Definition, es una especificación formal de los elementos estructurales

y de las definiciones necesarias para la codificación de ciertos tipos de documentos en SGML (véase la definición más abajo). Entre los DTD se incluyen MARC y TEI (véase la definición más abajo).

Dublin Core

Un conjunto mínimo de metadatos que los catalogadores, o los que crean documentos digitales, pueden asignar a los recursos de información y que pueden usarse para descubrir esos recursos, especialmente en la World Wide Web.

EAD

Encoded Archival Description, un DTD SGML que representa una forma muy estructurada de crear ayudas para la búsqueda digital en una colección de materiales de archivo.

elemento [element]

Un componente diferenciado de datos o metadatos. En algunos casos, los elementos contienen subcomponentes o subelementos, e.g. un elemento de dato "lengua" puede tener los subelementos "lengua usada en el registro de catalogación" y "lengua del material descrito en el registro de catalogación".

estándar de contenido de los datos [data content standard]

Las reglas o directrices que rigen el orden, la sintaxis y la forma en la que se introducen los valores de dato dentro de una estructura de dato, e.g., *Anglo-American Cataloging Rules* o *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries*.

estándar de estructura de datos [data structure standard]

Un estándar que define las categorías en las que se tiene que dividir la información. Este tipo de estándar especifica lo que constituye un registro y define las categorías de información o campos de datos, así como las

relaciones entre ellos. Algunos ejemplos son las *Categorías para la descripción de obras de arte*, el formato MARC y las Visual Resources Association Core Categories.

estándar de valor de datos [data value standard]

Los tesauros, fichas de autoridad, léxicos o listas de palabras que determinan qué términos se introducen en una estructura de datos, e.g., *Art & Architecture Thesaurus*, *Library of Congress Subject Headings*, *ICONCLASS* o *Union List of Artist Names*.

etiquetas [tags]

Breves nemotécnicas formales para indicar elementos de los datos o de los metadatos, especialmente en código HTML y SGML (e.g., <TITLE>, <META>).

hiper enlace [hyperlink]

Una abreviatura para "hypertext link". Los hiperenlaces, indicados por etiquetas codificadas, permiten determinar modos no lineales de desplazamiento entre o alrededor de documentos digitales o bien pueden enlazar objetos relacionados tales como imágenes o registros de audio.

hipermedia [hypermedia]

Una técnica para enlazar información de diversos tipos, frecuentemente de modo no lineal. Los enlaces están insertados en el texto, o en otro medio, mediante la introducción de etiquetas codificadas que el usuario no puede ver. Generalmente, se advierte al usuario de la existencia de un enlace por medio de texto subrayado. Cuando el usuario apunta a un enlace y lo selecciona, se activa el enlace y aparece la información asociada.

hostname

El identificador de una máquina en el Internet. El hostname no sólo identifica la máquina sino también su subnet y el dominio.

HTML

HyperText Markup Language, una lengua "markup" (véase la definición más abajo) derivada del SGML que se usa para crear documentos para las aplicaciones de la World Wide Web. El HTML enfatiza el diseño

más que la representación de la estructura del documento o de los elementos de los datos.

HTTP

HyperText Transfer Protocol, un protocolo estándar que permite a los usuarios con navegadores acceder a documentos HTML y a medios externos.

Internet

Una colección global de redes de ordenadores que intercambian información por medio del conjunto TCP/IP de protocolos de interconexión. Véase <http://www.fnc.gov/Internet-res.html>.

ISP

Internet Service Provider, una organización comercial que da acceso al Internet.

lengua markup [markup language]

Un modo formal de anotar un documento o una colección de datos digitales usando etiquetas de codificación insertadas específicamente para indicar la estructura del documento, o del fichero de datos, y el contenido de los elementos de los datos. El markup también da al ordenador información sobre cómo procesar y visualizar documentos ya anotados.

MARC

MACHINE-Readable Cataloging, un conjunto de estructuras de datos estandarizadas que se usan para describir los materiales bibliográficos que facilitan la catalogación cooperativa y el intercambio de datos en sistemas de información bibliográfica.

mecanismo de búsqueda [search engine]

Un programa que permite a los usuarios buscar en un sistema de información. Los mecanismos de búsqueda pueden permitir a los usuarios llevar a cabo la búsqueda de distintas maneras.

mecanismo de búsqueda del Internet (spider, crawler, robot) [Internet search engine]

Un programa de software que obtiene información de los metadatos y del contenido de las fichas que se encuentran en el Internet y

la pone en una base de datos dentro de la que los usuarios pueden buscar siguiendo diversos métodos. Los resultados de la búsqueda proporcionan enlaces con el sitio original de las fichas que coinciden con la búsqueda del usuario.

metadatos [metadata]

Literalmente, "datos sobre datos", los metadatos incluyen datos asociados tanto con un sistema de información como con un objeto informático para fines descriptivos, administrativos, legales, técnicos, de uso y conservación.

metadatos administrativos [administrative metadata]

Metadatos que se usan en la gestión y la administración de recursos de información, p. ej., información sobre la localización o el donante.

metadatos descriptivos [descriptive metadata]

Metadatos que se usan para describir o identificar recursos de información, e.g., registros de catalogación, URN (véase la definición más abajo) e índices especializados.

metadatos de conservación [preservation metadata]

Metadatos relacionados con la gestión de la conservación de los recursos de información, e.g., metadatos usados para documentar, o creados como resultado de los procesos de conservación practicados sobre los recursos de información.

metadatos de encabezamiento [header metadata]

Metadatos introducidos en la parte inicial de una ficha digital con el propósito de describirla o por cuestiones relacionadas con su gestión.

metadatos de uso [use metadata]

Metadatos, generalmente automáticamente creados por un ordenador, que están relacionados con el nivel y el tipo de uso de un sistema de información.

metadatos legales [legal requirements metadata]

Metadatos que documentan o siguen las condiciones asociadas con el uso o el acceso a recursos de información, e.g., privacidad y acceso o derechos y restricciones sobre la reproducción.

metadatos técnicos [technical metadata]

Metadatos creados para, o generados por, un sistema de ordenadores que están relacionados con el modo en que el sistema o su contenido se comportan o necesitan ser procesados.

MIME (Multipurpose Internet Mail Extension)

Un conjunto de especificaciones que hacen posible tanto el intercambio de texto entre lenguas con distintos caracteres como de correo electrónico multimedia entre distintos sistemas de ordenadores con estándares de correo del Internet.

modelo de contenido [content model]

Un modelo que define estructuras de datos y metadatos, incluyendo los tipos de elementos y subelementos que contienen.

multimedia [multimedia]

Materiales digitales, documentos o productos, como las páginas de la World Wide Web, los CD-ROM o los componentes de bibliotecas digitales y de museos virtuales que usan cualquier combinación de texto, datos numéricos, imágenes, animación, sonido o gráficos.

nesting

El modo en que los subelementos pueden estar contenidos dentro de elementos más grandes, constituyendo así múltiples niveles de metadatos.

objeto informático [information object]

Un ítem o conjunto de ítems, cualquiera que sea el tipo o el formato, que un ordenador puede usar o manipular como si se tratara de un objeto único.

"on the fly"

Una expresión que se usa para describir acciones que no han estado previamente definidas en el ordenador sino que ocurren según las necesidades del momento, p. ej., la generación de un conjunto de datos siguiendo las preferencias personales del usuario, o la conversión por solicitud de material codificado en SGML a HTML para ser presentado a un usuario que no tenga las aplicaciones necesarias para SGML.

protocolo [protocol]

Una especificación—a menudo estándar—que describe el modo en que se comunicarán los ordenadores entre sí.

protocolo para la transferencia de fichas (FTP) [file transfer protocol]

Un método para mover o transferir fichas entre ordenadores en el Internet.

SGML

Standard Generalized Markup Language, un estándar del ISO (International Organization for Standardization) que sirve para definir, especificar y crear documentos digitales que puedan ser enviados, visualizados, enlazados y manipulados de forma independiente del sistema.

spamming

(usado en referencia a metaetiquetas de "palabra clave"). Los metadatos que de forma abusiva se incluyen en el encabezamiento de las páginas de la Web para incrementar el número de visitantes a un sitio. El spamming comporta la repetición múltiple de una palabra clave para que así aparezca en la parte superior de la lista de resultados de un mecanismo de búsqueda; puede consistir también en introducir palabras clave que son irrelevantes para un sitio determinado con el fin de atraer visitantes bajo falsos pretextos.

TCP/IP

Transmission Control Protocol/Internet Protocol, el conjunto de protocolos de interconexión estandarizados del ISO que facilitan el enlace de sistemas de información institucionales a otros sistemas de información en el Internet, independientemente de

su plataforma de ordenadores. El TCP y el IP son dos estándares de comunicación para software que permiten la comunicación entre múltiples ordenadores en un contexto sin errores.

TEI

Text Encoding Initiative, una iniciativa colaborativa internacional encaminada a crear directrices genéricas para un esquema de codificación estándar de textos académicos.

URL

Uniform Resource Locator (también llamado Universal Resource Locator), una dirección del Internet que indica al usuario cómo y dónde encontrar una ficha determinada en la World Wide Web. Un URL no sólo incluye el nombre del fichero sino también el nombre del ordenador host, el directorio y el protocolo necesario (p. ej., <http://www.gii.getty.edu/aat-browser/>).

URN

Uniform Resource Name (llamado también Universal Resource Name/Number), el identificador único de una ficha del Internet que es independiente de la localización de la misma. La ficha continúa siendo accesible gracias a su URN, sin importar los cambios que puedan ocurrir en su host o en el recorrido del directorio.

valores por defecto [default values]

Valores suministrados automáticamente por un sistema de ordenadores cuando no se especifica un valor alternativo.

Web host

Web "hosting" se refiere al almacenamiento de un sitio o de una página en un servidor para hacerlos accesibles a través de la World Wide Web. La calidad de los servicios de Web hosting determina la eficacia de un recurso en el Internet. La calidad de un servicio de Web hosting se define de acuerdo con distintas consideraciones: la velocidad con que el servidor de la Web conecta con el Internet, el tipo de hardware y software que usa el servidor y el tipo de servicios avanzados (como el CGI scripting) que ofrece.

WHOIS++

Un estándar para un protocolo de servicios de directorio del Internet.

World Wide Web

Una arquitectura de cliente-servidor para acceder a documentos en hipermedia a través del Internet. La World Wide Web también ofrece medios para buscar fuentes remotas de información tales como bibliografías, listas telefónicas y manuales de instrucciones. Véase http://www.w3.org/MarkUp/HTMLPlus/htmlplus_3.html.

XML

Extensible Markup Language, un subconjunto simplificado de SGML, designado específicamente para ser usado con la World

Wide Web, que proporciona estructuras de datos más sofisticadas que el HTML.

Z39.50

Un protocolo estándar para la localización de información del ANSI (American National Standards Institute) que permite someter una busca a una base de datos independiente del tipo de hardware o software que ésta use. Originalmente implementado en el mundo bibliotecario, el Z39.50 se está usando ahora internacionalmente para la localización de información general.

véase también

Glossary of Networking Terms
(<http://www.cis.ohio-state.edu/htbin/rfc/rfc1208.html>)

NetGlos—The Multilingual Glossary of Internet Terminology
(<http://wwli.com/translation/netglos/>)

WWW Glossary
(<http://www.eelab.newpaltz.edu/~hill0/www/wwwgloss.html>)

Glosario recopilado por Murtha Baca, Tony Gill, Anne Gilliland-Swetland y Christina Yamanaka

Acrónimos usados en el texto y una selección de direcciones de la World Wide Web

AACR2

Anglo-American Cataloguing Rules
(2nd edition)

AAH

Association of Art Historians
<http://scorpio.gold.ac.uk.aaah/>

AAT

Art & Architecture Thesaurus
<http://www.gii.getty.edu/vocabulary/aat.html>

ACN

Advisory Committee on Networking
<http://www.jisc.ac.uk/acn/index.html>

ADAM

Art, Design, Architecture & Media Information Gateway
<http://adam.ac.uk/>

AGOCG

Advisory Group on Computer Graphics
<http://www.agocg.ac.uk/>

AHDS

Arts & Humanities Data Service
<http://www.ahds.ac.uk/>

AMICO

Art Museum Image Consortium
<http://www.amn.org/AMICO/>

ANR

Access to Network Resources
<http://ukoln.ac.uk/elib/lists/anr.html>

ARLIS

Art Libraries Society
<http://www.lib.duke.edu/lilly/arlis/index.html>

BHA

Bibliography of the History of Art
www.gii.getty.edu/bha/index.html

Biz/Ed

Business Education on the Internet
<http://www.bizednet.bris.ac.uk:8080/>

CAIN

Conflict Archive on the Internet
<http://cain.ulst.ac.uk/>

CALT

Committee on Awareness, Liaison and Training
<http://www.jisc.ac.uk/>

CDWA

Categories for the Description of Works of Art
<http://www.gii.getty.edu/cdwa>
CDWA cataloging examples:
<http://www.ahip.getty.edu/cdwa/examples/home3.htm>

CEI

Committee for Electronic Information
<http://www.jisc.ac.uk/cei>

CHArt

Computers and the History of Art
<http://giorgio.hart.bbk.ac.uk/chart/>

CHIN

Canadian Heritage Information Network
<http://www.pch.gc.ca>

CIMI

Computer Interchange of Museum Information
<http://www.cimi.org>
The CIMI Schema:
<http://www.cimi.org/downloads/ProfileFinalMar98/cimiprofile4.htm#6.4.3.2>

CTI

Computers in Teaching Initiative
<http://info.ox.ac.uk/cti/>

DC

Dublin Core
<http://www.oclc.org/oclc/research/projects/core/index.htm>

DDC

Dewey Decimal System
<http://www.oclc.org/oclc/ftp/index.htm>

DHS

Design History Society
<http://www.sequence.co.uk/dhs/June97>

ECSTASY

Enhanced Collaboration with Shared Tools for Art and Design Systems
<http://www.rave.ac.uk/ecstasy/>

EEVL

Edinburgh Engineering Virtual Library
<http://eevl.icbl.hw.ac.uk/>

eLib

Electronic Libraries Programme
<http://www.ukoln.ac.uk/services/elib/>

FDA

Foundation for Documents of Architecture: *A Guide to the Description of Architectural Drawings*
<http://www.gii.getty.edu/fda>

GII

Getty Information Institute
<http://www.gii.getty.edu>

GLADNet

Group for Learning in Art & Design Network
<http://www.mailbase.ac.uk/lists-f-j/gladnet/>

HEDS

Higher Education Digitisation Service
<http://heds.herts.ac.uk/>

HEFCs

Higher Education Funding Councils
<http://www.niss.ac.uk/education/hefc/index.html>

HTML

HyperText Markup Language

ICOM

International Council of Museums
<http://www.icom.org>

IHR-Info

Institute for Historical Research Information Gateway
<http://ihr.sas.ac.uk/>

ISO

International Organization for Standardization
<http://www.iso.ch/>

JISC

Joint Information Systems Committee
<http://www.jisc.ac.uk/>

MDA

Museum Documentation Association
<http://www.open.gov.uk/mdocasn/index.htm#MDAHomePage>

MESL

Museum Educational Site Licensing Project
<http://www.gii.getty.edu/mesl/>

MODELS

Moving to a Distributed Environment for
 Library Systems
<http://www.ukoln.ac.uk/models/intro.html>

NPO

National Preservation Office
<http://www.bl.uk/services/preservation/>

Object ID

<http://www.gii.getty.edu/pco>

OCLC

Online Computer Library Center
<http://www.oclc.org/>

OMNI

Organising Medical Networked Information
<http://omni.ac.uk/>

REACH

Record Export for Art and Cultural Heritage
<http://www.rlg.org/reach.html>

RLG

The Research Libraries Group
<http://www.rlg.org>

ROADS

Resource Organisation and Discovery
 System
<http://ukoln.ac.uk/roads/>

SCRAN

Scottish Cultural Resources Access Network
<http://www.scran.ac.uk/>

SOSIG

Social Science Information Gateway
<http://sosig.ac.uk/>

SuperJANET

Super Joint Academic Network
<http://www.ja.net/>

TASC

Technology Applications Sub-Committee
<http://www.jisc.ac.uk/tasc/index.html>

TGN

Getty Thesaurus of Geographic Names
<http://www.gii.getty.edu/vocabulary/tgn.html>

TLTP

Teaching and Learning Technology
 Programme
<http://www.tltp.ac.uk/>

TLTSN

Teaching and Learning Technology Support
 Network
<http://www.tltp.ac.uk/tltsn/>

UKOLN

United Kingdom Office for Library and
 Information Networking
<http://ukoln.bath.ac.uk/>

ULAN

Union List of Artist Names
<http://www.gii.getty.edu/vocabulary/ulan.html>

VADS

Visual Arts Data Service
<http://vads.ahds.ac.uk>

VISION

Visual Resources Sharing Information
 Online Network
<http://www.rlg.org/pr/9711vis.html>
<http://www.oberlin.edu/~art/vra/vision.html>

VRA

Visual Resources Association
<http://www.oberlin.edu/~art/vra/vra.html>
 VRA Core Categories:
<http://www.oberlin.edu/~art/vra/wc1.html>

WHOIS++

<http://www.nlc-bnc.ca/pubs/netnotes/notes54.htm>

W3C

World Wide Web Consortium
<http://www.w3.org/>

Z39.50

<http://www.ukoln.ac.uk/dlis/z3950/defin.html>

Colaboradores

Willy Cromwell-Kessler

(bl.kes@rlg.org) se integró al Research Libraries Group (RLG) en 1997, tras haber pasado varios años en las bibliotecas de la Stanford University, donde ejerció diversos cargos. Como Head of Cataloging y Coordinator for Cataloging Development, se ocupó de integrar en el catálogo en línea de la biblioteca de la universidad recursos de información digitales, y otros recursos no tradicionales. Actualmente, es la Bibliographic Specialist en una nueva división del RLG, la Integrated Information Services Division, donde junto con sus colegas está trabajando en el desarrollo de datos estructurados para una futura red del RLG que incluirá texto íntegro digital, bibliografía, citas y otros bancos de datos. Ha trabajado también en la creación de estándares para bibliotecas y bibliográficos, y ha formado parte de grupos dedicados al establecimiento de estándares. Su participación más reciente ha sido como Chair of the Standards Committee en el Program for Cooperative Cataloging (PCC), un grupo que ha desarrollado diversos estándares esenciales de registro PCC. Willy Cromwell-Kessler obtuvo un M.A. en literatura inglesa y un M.L.S. por la Indiana University en Bloomington.

Tony Gill (tony@adam.ac.uk) ocupa el cargo de Programme Leader para los proyectos ADAM y VADS en el Surrey Institute of Art and Design, y se responsabiliza de la implementación del Art, Design, Architecture & Media (ADAM) Information Gateway y del Visual Arts Data Service (VADS), cuyo objetivo es facilitar el acceso a información interconectada de alta calidad sobre las artes visuales. Anteriormente, ejerció la función de Technical Outreach Manager en el Museum Documentation Association (MDA), donde asesoraba a museos y galerías sobre el uso de tecnologías de la

información. Ha obtenido títulos universitarios en Communication in Computing (Middlesex University) y en Physics and Philosophy (King's College, Londres), y actualmente es un estudiante de doctorado en la School of Information Systems de la Kingston University. Ha publicado diversas obras sobre la aplicación de las tecnologías de la información en el campo de las artes y de las humanidades, entre ellos la *MDA Guide to Computers in Museums*. Tony Gill ha participado activamente en cuatro de los cinco talleres del Dublin Core, ha presentado varios trabajos relacionados con el tema de los metadatos, y ha escrito (en colaboración con Paul Miller) dos de los documentos "Request for Comment" del Dublin Core. Forma parte de los comités del CHART (Computers and the History of Art), del Joint Information Systems Committee Image Digitisation Initiative Implementation Group, y del Technical Advisory Service for Images Steering Group.

Anne J. Gilliland-Swetland

(swetland@ucla.edu) es Assistant Professor del Department of Library and Information Science de la Graduate School of Education and Information Studies de la University of California en Los Angeles. Su campo de especialización en la investigación y en la enseñanza incluye los métodos de archivo, la gestión de registros electrónicos, y la administración de información digital. Actualmente está llevando a cabo una evaluación del uso del Dublin Core, como participante en un proyecto llamado Digital Portfolio Archives in Elementary Science Learning que ha financiado la National Science Foundation; también está estudiando la aplicación del Encoded Archival Description (EAD) en la California Digital Library. Obtuvo un M.A. en English Language and Literature por el Trinity College de la University of Dublin; un M.S. y un C.A.S. en

Library and Information Science por la University of Illinois en Urbana-Champaign; y un Ph.D. en Information and Library Studies por la University of Michigan. En su tesis doctoral investigó la posibilidad de crear un área de especialización para la valoración de las comunicaciones electrónicas. La Dra. Gilliland-Swetland ha publicado numerosos trabajos educativos y de investigación sobre la aplicación de tecnologías digitales en museos y en archivos. Actualmente está escribiendo un libro (en colaboración con Gregory Leazer) titulado *Content, Context and Structure in the Organization of Recorded Knowledge*, en el que se examinará el desarrollo de métodos para la organización de la información en varios grupos profesionales. Es miembro del EAD Working Group y del Council of the Society of American Archivists.

Murtha Baca (mbaca@getty.edu) obtuvo un Ph.D. en Art History and Italian Literature por la University of California en Los Angeles, y actualmente ocupa el cargo de Project Manager for Standards Initiatives en el Getty. Entre sus publicaciones se incluyen traducciones del italiano de numerosas monografías y catálogos de exhibiciones, seis manuscritos de Leonardo da Vinci, *An Italian Renaissance Sextet: Six Tales in Historical Context* (Marsilio Publishers, 1994) y *Science in the Kitchen and the Art of Eating Well* de Pellegrino Artusi (Marsilio Publishers, 1997). En colaboración con James M. Bower, editó la *Union List of Artist Names* (G. K. Hall, 1994). Entre sus artículos más recientes se incluyen "From Authority File to Retrieval Tool: The Union List of Artist Names" (*Computers and the History of Art*, vol. 6, no. 2, 1997) y "Making Sense of the Tower of Babel: A Demonstration Project in Multilingual Equivalency Work" (*Terminology: International Journal of Theoretical and Applied Issues in*

Specialized Communication, 1997). Se ocupó de supervisar la producción y la publicación de *Categories for the Description of Works of Art*, un documento impreso y en hipertexto, y editó, junto con Patricia Harpring, un número doble especial de la revista *Visual Resources* (vol. XI, nos. 3–4,

1996). Es la representante del Getty en el International Terminology Working Group, entre cuyos proyectos colaborativos se incluye un léxico multilingüe de objetos litúrgicos. La Dra. Baca supervisó la publicación en cuatro lenguas de *Guidelines for Forming Language Equivalents:*

A Model Based on the Art & Architecture Thesaurus (Getty Information Institute, 1996), y actualmente está coordinando un proyecto con el Ministerio de Cultura de Chile cuyo propósito es proporcionar equivalentes en español para varias jerarquías del *Art & Architecture Thesaurus*.

El continuo crecimiento de la información digital asequible a través de las redes electrónicas ha creado la necesidad apremiante de desarrollar estándares comunes que faciliten la localización, la recuperación y la gestión de esta compleja masa de información. Los metadatos, que podrían ser definidos como "información sobre información", proporcionan un modo de indexar, acceder y encontrar los recursos digitales.

En *Introducción a los metadatos* se discuten los metadatos desde distintas perspectivas, con un énfasis en los sistemas de metadatos para información sobre el patrimonio cultural. Este libro, útil para todos aquellos interesados en el acceso a la información electrónica, presta especial atención a los problemas que emergen al tratar de crear metadatos y explora varias iniciativas en proceso de desarrollo.

La serie *Introducción a . . .* familiariza a profesionales e investigadores con las complejas cuestiones y tecnologías relacionadas con la producción, gestión y difusión de los recursos de información sobre el patrimonio cultural.

