



World Library and Information Congress: 69th IFLA General Conference and Council

1-9 August 2003, Berlin

Code Number: 053-S
Meeting: 157. Bibliographic Control - **Workshop**
Simultaneous Interpretation: -

Modelos de Metadatos para Pasarelas Temáticas

Lynne C. Howarth

Facultad de Estudios de la Información
Universidad de Toronto
Toronto, Canadá
howarth@fis.utoronto.ca

Resumen: Las pasarelas temáticas habilitadas para la Web brindan acceso a colecciones de recursos de alta calidad en un área temática específica. Las descripciones de documentos, objetos o sitios cuidadosamente seleccionados y evaluados identifican unívocamente el contenido digital y enlazan con él. Según analizará esta ponencia, la determinación del objetivo y del contenido correspondiente para una pasarela temática influirá necesariamente en la selección subsiguiente de *estructuras* y *modelos* de metadatos. También se evaluarán otras consideraciones relativas al nivel de detalle e interoperatividad de los metadatos. La exposición concluirá con un análisis de los desafíos y las tendencias futuras con respecto al uso de modelos de metadatos en las pasarelas temáticas.

1.0 Introducción

Como lo confirma una búsqueda simple en Google™, las definiciones del término “pasarelas temáticas” (subject gateways) son casi tan numerosas como los servicios que describen. Reconociendo, sin embargo, la necesidad de un punto de partida, Traugott Koch (2000, 2) propone lo siguiente como un intento de definir el concepto:

Las pasarelas temáticas son servicios de Internet que brindan apoyo a la búsqueda sistemática de recursos. Proporcionan enlaces a recursos (documentos, objetos, sedes de servicios), predominantemente accesibles a través de Internet. El servicio se basa en descripciones de recursos. El acceso de consulta a los recursos a través de una estructura temática es una característica importante.

Koch (2000, 1) observa además que, “Se emplea un considerable esfuerzo manual para garantizar una selección de recursos que reúnan criterios de calidad y muestren una descripción rica de estos recursos con metadatos basados en normas.” Con esta explicación adicional del proceso, se clarifica la conexión entre los conceptos de “pasarelas temáticas” y “modelos de metadatos” y se establecen los fundamentos para explorar cómo estos últimos pueden utilizarse con máxima efectividad en la creación y soporte de las primeras. También se expondrán en las siguientes secciones de la ponencia los desafíos específicos y tendencias futuras asociados a aplicaciones de metadatos en pasarelas temáticas.

2.0 Consideraciones y Aplicaciones

En general, puede hacerse una distinción entre metadatos de formato simple – tales como los representados en la sintaxis de un lenguaje de marcas (ej., XML; HTML; SGML), e insertados dentro de la estructura del objeto digital – y metadatos de formato muy estructurado. Para los primeros, los robots de búsqueda (web crawlers) o los “bots”, pueden recuperar las metaetiquetas especificadas (ej., <Título>) a fin de extraer valores determinados y, con una mínima (o ninguna) intervención humana, disponer los resultados dentro de índices invertidos o directorios de temas predefinidos. Las categorías temáticas aplicadas en los sitios de motores de búsqueda Yahoo!™, AltaVista™ o Google™ son ejemplos bien conocidos. Los problemas que pueden surgir con los vocabularios del lenguaje natural, el texto no estructurado y con lo que la comunidad del control bibliográfico reconoce como la falta de “control de autoridades”, son endémicos en el ámbito de los metadatos de “formato simple”. Sin embargo, la resistencia a la estructura – ya sea impuesta o adoptada voluntariamente – en el Internet libertario continúa fomentando iniciativas del usuario final, tales como el Open Directory Project (Proyecto de Directorio Abierto), donde los creadores de recursos pueden asignar la terminología temática que han ideado al margen de cualquier norma (véase: <http://www.dmoz.org>)

En cambio, los metadatos de formato muy estructurado se idean, aplican y mantienen de acuerdo con normas (internacionales) claramente establecidas. Estos son los modelos de metadatos formales que se utilizan en pasarelas temáticas “de calidad controlada” (Koch 2000) y que proporcionan un valor añadido a la búsqueda de recursos. Al crear una pasarela temática habilitada para metadatos “de calidad controlada”, deben considerarse tres aspectos clave, a saber: qué recursos incluir (documentos; objetos; sitios), qué tipo de estructuras de metadatos emplear para describir y acceder a esos recursos y qué modelo de metadatos aplicar al crear los registros para enlazar con el contenido. Cada uno de estos aspectos se examinará, respectivamente, en las siguientes secciones.

2.1 *Cómo Determinar el Objetivo y el Contenido para Pasarelas Temáticas*

La determinación del objetivo y el contenido correspondiente para una pasarela temática influirá necesariamente en la posterior selección de las *estructuras* y *modelos* de metadatos. ¿Cuál

Tabla 1
Criterios de Inclusión para Determinar el
Alcance y el Contenido de una Pasarela Temática

Criterios de Inclusión	Único (ejemplos)	Múltiple (ejemplos)	Universal (ejemplos)
Materia/Tema	Leucemia	Enfermedades Cardiacas y Neurológicas	Todas las enfermedades
Lengua	Inglés	Alemán, Francés, Japonés, Griego	Todas las lenguas
Localización Geográfica	Canadá	Europa y Asia	Todos los países, regiones, etc.
Período de tiempo	2003	1900-1999	Toda la historia registrada
Tipo de recurso	Sitios de la Web	Sitios de la Web, depósitos de datos y archivos fotográficos	Todos los recursos habilitados para la Web
Grupos/Asociaciones	Mujeres	Niños y adultos jóvenes	La humanidad
Formato del material	Texto electrónico	Documentos Word, mapas digitales, DVD	Todos los formatos de texto y soportes (analógicos y digitales)

es la intención específica de la pasarela temática, qué objetivos concretos han de lograrse en su diseño y qué producto final o resultados se esperan de ella? El alcance y la cobertura, las delimitaciones y límites de la pasarela temática implican criterios tales como materia o tema, lengua, localización geográfica, periodo de tiempo, tipo de recurso, grupos o asociaciones, formato del material, etc. La Tabla 1 resume, con ejemplos, cómo podría combinarse y evaluarse cada criterio a fin de determinar el diseño final de la pasarela temática. Obsérvese que podría adoptarse cualquier número de enfoques y que el siguiente ofrece sólo una muestra, cuando podrían aplicarse muchas permutaciones o combinaciones. La definición de “uno”, “muchos” o “todos” de cualquiera de los criterios enumerados en la primera columna también es relativa, y claramente abierta a diversas interpretaciones.

Una vez que se han tratado estas cuestiones, pueden determinarse criterios para la selección de recursos y las fuentes u objetivos para aquellos recursos identificados. Koch (2000, 6) sugiere que los modelos que se dan con más frecuencia para las pasarelas temáticas incluyen:

- Específico de un tema nacional (una materia; un país; una lengua – ej., GEM)
- Temas cruzados nacionales (materias múltiples; un país; una lengua – ej., DutchESS)
- Específico de un tema global (una materia; global; una lengua – ej., EEVL)
- Temas cruzados globales (materias múltiples; global; una lengua – ej., ADAM)
- Universal (todas las materias; global; varias lenguas – ej., CORC)

Podría añadirse que la tipología precedente no es ni definitiva ni exhaustiva y , como la Tabla 1, añade una visión o perspectiva más de las posibles formas de combinar criterios

de selección para diseñar pasarelas temáticas basadas en un determinado enfoque o en objetivos específicos por lograr.

2.2 *Cómo Determinar las Estructuras de Metadatos Aplicables*

Como se ha señalado previamente, la determinación del alcance, la cobertura y el contenido derivado de los criterios de selección de una pasarela temática influirá sobre qué elementos y modelos de metadatos se elegirán para dar soporte a la identificación, y sus enlaces, de los recursos señalados. En general, los *tipos* o *estructuras* de metadatos que podrían requerirse para dar soporte a una pasarela temática, confeccionada a partir de la configuración de los criterios de selección esbozados en la Tabla 1, incluyen los siguientes:

- Metadatos administrativos: información de “mantenimiento” sobre el propio registro – su creación, modificación, relación con otros registros, etc. Los ejemplos de elementos que pertenecen a metadatos administrativos incluyen los siguientes, sin limitarse a ellos:
 - Número de registro
 - Fecha de creación del registro
 - Fecha de la última modificación
 - Identificación del creador / revisor del registro
 - Lengua del registro
 - Notas
 - Relación de este registro con otro(s)
- Metadatos descriptivos: describen las propiedades físicas e intelectuales o el contenido de un ítem u objeto digital con elementos tales como:
 - Título (también títulos alternativos y paralelos; subtítulos; títulos abreviados; etc.)
 - Creador (autor; compositor; cartógrafo; artista; etc.)
 - Fecha
 - Editor
 - Identificadores únicos y enlaces dinámicos (URI; URL; etc.)
 - Resumen, nota descriptiva; revisión; etc.
 - Nivel de audiencia
 - Soporte físico; formato; etc.
- Metadatos analíticos: información que analiza y facilita el acceso a los contenido del recurso. Citados a veces como “metadatos temáticos”, los elementos pueden incluir:
 - Encabezamientos de materia
 - Tesauros
 - Palabras clave de materia / tema
 - Resumen; Índices (TOC)
 - Códigos de clasificación derivados de sistemas de clasificación
 - Otros elementos de importancia local, ej., afiliación departamental; enlace con otro contenido electrónico relacionado; etc.
- Metadatos de gestión de derechos: información relativa a las restricciones (legales; económicas; etc.) sobre el acceso a ítems u objetos digitales o su utilización. Pueden aplicarse elementos tales como los siguientes:

- Restricciones sobre el uso
- Declaraciones de permisos
- Tasas de suscriptor / licencia / pagos por uso
- Reconocimientos
- Nota de derechos de autor
- Plazos de retención
- Calificaciones de calidad
- Advertencias de uso
- Metadatos técnicos: el equipo o programa informático concreto utilizado para convertir un ítem / objeto a un formato digital, o al almacenarlo, visualizarlo, etc., puede requerir el uso de elementos tales como:
 - Especificaciones del equipo de digitalización
 - Posiciones de la cámara
 - Condiciones de filmación
 - Parámetros de codificación
 - Equipo y programa informático de reconocimiento de voz y/o reproducción
 - Especificaciones del escáner óptico
 - Equipo de generación de imágenes
 - Tipo de fichero y requisitos del programa informático de conversión
- Otros, según se determine – ej., elementos específicos de metadatos basados en requisitos locales, regionales, organizativos, o de acuerdo con una norma de metadatos establecida a nivel nacional, y no incluida en los tipos de metadatos citados anteriormente.

2.3 *Cómo Seleccionar uno o más Modelos de Metadatos*

La elección de uno o más modelos de metadatos a emplear en la creación de los registros secundarios para identificar y enlazar unívocamente con los recursos accesibles a través de la pasarela temática dependerá de la intención concreta del servicio y de los tipos de metadatos que contenga. De esta manera, una pasarela temática creada y mantenida por una red distribuida de organizaciones nacionales que contenga sitios Web de alta calidad (solamente texto e imágenes), y limitada a un área temática en un dominio técnico, podría requerir una mezcla de metadatos administrativos, descriptivos y analíticos. La Red Sanitaria Canadiense ofrece un ejemplo de tal configuración. Una “exposición virtual” que contiene enlaces a diversos objetos digitales contenidos dentro de un consorcio internacional de galerías de arte y museos, públicos y privados, necesitaría el uso de metadatos técnicos y de gestión de derechos, además de los requeridos para fines administrativos, descriptivos y analíticos.

Lo que puede ayudar a la determinación final del modelo de metadatos es el grado deseado de *minuciosidad*, o sea, la cantidad de detalles a recoger y representar en el registro de metadatos. Un “registro base” – creado utilizando un modelo de metadatos como el Dublin Core con su juego de quince elementos (todos ellos optativos, repetibles y extensibles) – recoge perfectamente los metadatos administrativos, descriptivos, analíticos y de gestión de derechos, y puede incorporar información vinculada con especificaciones técnicas. Sin embargo, en algunos ámbitos especializados, un modelo de metadatos como el Dublin Core, carece de minuciosidad (nivel de detalle) suficiente

para representar adecuadamente los recursos, o los fines concretos a los que se orienta la pasarela temática. El estándar de metadatos ONIX para la edición y los editores internacionales, o el Estándar de Contenidos para Metadatos Geoespaciales Digitales son dos ejemplos de modelos de metadatos muy detallados y sumamente técnicos, desarrollados especialmente para gestionar contenidos complejos y aplicaciones únicas dentro del ámbito.

Además de decidir el nivel de detalle a recoger en los registros habilitados para metadatos, la elección del modelo puede afinarse según la respuesta a preguntas tales como las siguientes:

- ¿Está la pasarela temática propuesta en un dominio (tema o disciplina) para el cual se ha desarrollado un estándar de metadatos de formato muy estructurado?
- ¿Qué campos serían más útiles para la comunidad de usuarios a los que la pasarela temática pretende servir? ¿Qué nivel de detalle deberían recoger esos campos?
- ¿Qué campos serían más útiles para quienes crean y/o mantienen la pasarela temática? ¿Qué nivel de detalle deberían recoger esos campos?
- ¿Qué campos se requerirían para apoyar a servicios concretos que la pasarela temática pretende proporcionar?
- ¿El uso o acceso a la pasarela temática estará restringido de alguna manera? ¿Cómo se registrará (o debería registrarse) en los metadatos del registro?
- ¿Hay algunos requisitos vinculados a la lengua, al formato del material o al tipo de soporte para los que se deben aportar campos específicos (o adicionales)?
- ¿Hay requisitos para crear o compartir recursos entre una red de colaboradores con responsabilidad sobre la pasarela temática? ¿Se requieren campos de metadatos (adicionales) para la gestión de la pasarela?
- Si se prevé o se requiere el uso de más de un modelo de metadatos (compartiendo recursos a través de redes), ¿están disponibles, pronta e inmediatamente, los mapeos entre modelos cruzados autorizados (vías cruzadas) para facilitar y mantener la interoperatividad? ¿Pueden intercambiarse los recursos representados en un modelo (o estándar) de metadatos entre los colaboradores de la pasarela temática que estén utilizando un modelo (o estándar) distinto?
- ¿Cuánto se utiliza un modelo concreto y en qué aplicaciones o entornos comparables al propuesto actualmente? ¿Qué consistencia y/o flexibilidad tiene el modelo en diferentes contextos?
- ¿Con qué rapidez se puede migrar desde este modelo concreto hacia otro, se necesitará la conversión de datos en algún momento?
- ¿Cómo o con qué fidelidad se adapta un modelo concreto a los estándares establecidos a nivel organizativo (local), nacional o internacional, si los hubiera?
- ¿Qué recursos humanos (cantidad; formación; preparación), técnicos, económicos u otros se requieren para respaldar la aplicación del modelo de metadatos y tiene mi organización u operación esos recursos disponibles, con prontitud y en cantidad suficiente? ¿Hay otras restricciones prácticas para implementar y mantener uno o más modelos concretos?

Una vez respondidas algunas o varias de las preguntas precedentes, puede confirmarse la elección de emplear uno o más modelos de metadatos *normalizados*. Por otra parte, una persona, organización o consorcio que decida crear una pasarela temática

puede determinar cuál es la opción preferida, una solución local o “cosecha propia” – conjunto de elementos de metadatos localmente determinados y respaldados – . Igualmente, algunos deciden combinar elementos de un estándar establecido, como el Dublin Core, con elementos apropiados a la situación local de recursos y objetivos. No hay una receta única ni solución de “talla única” en cuanto a qué modelo o estándar de metadatos utilizar en una pasarela temática.

3.0 Desafíos Presentes y Tendencias Futuras

En general, los desafíos presentes son buenos vaticinadores de las cuestiones que requerirán atención especial en el futuro, ya sea a corto o largo plazo. A medida que se extiende el número, la cobertura, el alcance y las expectativas del usuario final de las pasarelas temáticas habilitadas para metadatos, persistirán los problemas de algunas áreas, como las siguientes, y susceptibles de solución:

- Interoperatividad – aumentará la exigencia de mejoras en los protocolos de metadatos de dominio cruzado y vías cruzadas, que posibiliten el intercambio de registros; será necesario desarrollar o mejorar estándares de metadatos que faciliten la interoperatividad a nivel técnico, semántico, organizativo, intercomunitario e internacional
- Colaboración y cooperación – las pasarelas temáticas pueden expandirse empleando economías de escala, para promover el acceso, habilitado para metadatos, a recursos internacionales en dominios cruzados, así como para compartir las oportunidades y los costes para crear y mantener en común los correspondientes modelos de metadatos
- Extensión – Si bien está claro que algunas pasarelas temáticas se autolimitan, parece inevitable el crecimiento adicional en los servicios definidos como de amplio alcance o inclusión (ej., CORC de OCLC; UKOLN). Si bien la concreción de una pasarela temática verdaderamente universal es improbable en un futuro inmediato, es un objetivo que debería preverse y enfrentarse a las carencias de las estructuras o elementos de metadatos
- Recursos multilingües – hasta ahora, se han desarrollado y aplicado modelos de metadatos en entornos monolingües; las demandas del usuario final para acceder a recursos multilingües en la lengua de su elección exigirán modelos de metadatos nuevos o significativamente amplios y mejorados, o bien aplicaciones innovadoras de modelos existentes no verbales de metadatos (ej., sistemas de clasificación) para describir y recuperar recursos multilingües
- Motor de búsqueda y funcionalidad de la interfaz – motores de búsqueda de sofisticación creciente explotarán de manera más efectiva toda la riqueza existente de metadatos para facilitar la localización de recursos; la búsqueda o recuperación avanzada entre pasarelas puede requerir el desarrollo de modelos nuevos o expandidos, tales como los metadatos de descripción de colecciones, los metadatos de productos o procesos, los metadatos para nuevos formatos o innovaciones técnicas, etc.; asimismo, pueden requerirse nuevos elementos para describir o apoyar una funcionalidad de interfaz mejorada (ej., accesibilidad; facilidad de uso; navegación; características para responder a necesidades especiales; etc.)

- Herramientas de desarrollo de metadatos – Hasta ahora, la creación de registros de metadatos se ha visto como un ejercicio eminentemente manual (y humano). A medida que prolifere el número de pasarelas temáticas, aumentará la aplicación de de tratamientos automáticos, aunque sólo sea como una primera intervención que requiera una posterior mediación humana. Las herramientas que recogen e indizan recursos automáticamente, llenando estructuras predefinidas de registros de metadatos, se emplearán crecientemente para la gestión de pasarelas temáticas
- Archivos y utilización local – las modificaciones y expansiones de los modelos de metadatos existentes han dado como resultado la creación de archivos para registrar y rastrear los cambios; el número de variaciones locales del estándar ha acelerado el crecimiento de archivos; en el futuro, puede haber una tendencia a desaconsejar o limitar las aplicaciones no normalizadas de carácter local, independientemente de la disponibilidad de un registro
- Políticas, cuestiones legales, autenticación – el crecimiento de las pasarelas temáticas puede hacer necesaria la expansión o la mejora de metadatos vinculados a la gestión de pasarelas, a los derechos de propiedad intelectual (IP) y a la originalidad/autenticidad de los recursos
- Cumplimentación de estándares – sera prioritario asegurar e incluso obligar a la cumplimentación de estándares o protocolos internacionales de metadatos de dominios cruzados, a media que proliferen las pasarelas temáticas y se comprometan más colaboradores

Como puede ilustrar lo precedente, la gama y diversidad de cuestiones que surgen de la aplicación de modelos de metadatos en las pasarelas temáticas son lo bastante numerosas como para comprometer a investigadores y profesionales de los metadatos, por igual, durante parte del futuro.

Bibliografía Selecta:

Campbell, Debbie (2000). *Australian subject gateways: metadata as an agent of change [Pasarelas temáticas australianas: los metadatos como un agente del cambio]*. Presentado en la Conferencia VALA 2000, Melbourne, Australia, 18 de febrero de 2000. 7 páginas. Disponible en: <http://www.nla.gov.au/nla/staffpaper/dcampbell2.html> Acceso del 10/05/03.

DESIRE Information Gateways Handbook [Manual de Pasarelas de Información DESIRE]. 2000. 149 páginas. Disponible en: <http://www.desire.org/handbook/print4.html> Acceso del 30/04/03

De Jong, Annamieke (2002). Dominio audiovisual. En *SCHEMAS Metadata Watch Report #8 and Standards Framework Report #4: Appendix A. [Informe de Observación #8 de Metadatos SCHEMAS e Informe de Marco #4 de Estándares: Apéndice A]* 6 páginas. Disponible en: <http://www.schemas-forum.org/metadata-watch/d29/d29.htm> Acceso del 24/4/03

Howarth, Lynne C. (2003). *Metadata unplugged [Los metadatos desentrañados]*. Una presentación para la Conferencia del 25º Aniversario de los Bibliotecarios y Especialistas

en Información por Abuso de Sustancia, Toronto, Canadá, 25 de abril de 2003.
Disponible en: <http://www.SALIS.org/> Acceso del 19/05/03

Koch, Traugott (2000). Pasarelas temáticas con calidad controlada: definición, tipologías, panorama empírico. *Online Information Review*. Vol. 24:1. 17 páginas.
Disponible en: <http://www.lub.lu.se/tk/publ/OIR-SBIG.html> Acceso del 07/05/03.