

# Pasarelas temáticas en Internet: un modelo de directorio basado en la aplicación de técnicas documentales

**Luís Rodríguez Yunta**

Citación recomendada: Luís Rodríguez Yunta. *Pasarelas temáticas en Internet: un modelo de directorio basado en la aplicación de técnicas documentales* [en línea]. "Hipertext.net", núm. 2, 2004. <<http://www.hipertext.net>> [Consulta: 12 feb. 2007]. .

1. Directorios y/o buscadores: dos sistemas de recuperación complementarios
2. Pasarelas temáticas y aplicación de las técnicas de análisis documental
3. Normalización y sistematización
  - 3.1. Normalización interna
  - 3.2. Normalización entre sistemas
4. Ejemplos de pasarelas temáticas
  - 4.1. Reino unido
  - 4.2. Resto de Europa
  - 4.3. Estados unidos
  - 4.4. Australia y nueva zelanda
  - 4.5. España
5. Problemática de la construcción de pasarelas temáticas
  - 5.1. Falta de acuerdo en cuanto a los campos de las bases de datos
  - 5.2. Dificultad en la definición de la unidad documental y la tipología de recursos
  - 5.3. Necesidad de fijar unos criterios de selección
  - 5.4. Preocupación por la usabilidad y recuperabilidad
  - 5.5. Inseguridad del mantenimiento
6. Notas

## 1. Directorios y/o buscadores: dos sistemas de recuperación complementarios

La búsqueda de información en Internet es hoy en día una tarea imprescindible para múltiples profesionales. Para guiarse a través de la red se dispone de una amplia oferta de herramientas. La opción más utilizada es la de los buscadores, bases de datos que permiten localizar páginas web por las palabras presentes en ellas. Para gran parte de las consultas, es el sistema de recuperación más rápido y eficaz. No llegan a poder interrogar toda la red, pero su campo de acción es enorme y se va ampliando progresivamente con una actualización muy aceptable, aunque sus resultados pueden ser muy aleatorios. Frecuentemente ofrecen tal volumen de respuestas que no resulta fácil discriminar lo esencial de lo anecdótico.

También se puede localizar mucha información navegando a través de las páginas de enlaces seleccionadas por una institución con el aval de su prestigio, o a través de una red de instituciones que formen un anillo (webring). Cuando estas recopilaciones de enlaces son especialmente ricas en relaciones, se constituyen en directorios y guías de Internet, sistemas de información clasificada por temas de interés. A menudo pueden quedar desfasados o resultar incompletos, pero permiten obtener unas respuestas más sistemáticas que las ofrecidas por un buscador y más completas que una sencilla página de enlaces.

Buscadores y directorios son opciones complementarias. Ambos trabajan sobre una base de datos y ofrecen la posibilidad de interrogar por palabras. Pero en el primer caso los resultados obtenidos se corresponden con páginas web, definidas por una dirección URL específica y única. Por el contrario en los directorios los resultados se corresponden generalmente con sedes web, definidas por un dominio que abarca múltiples páginas. Por ejemplo si se interroga por "Cindoc" en el buscador Google ( <http://www.google.es/> ) se obtienen miles de páginas web en las que aparece este término, tanto las presentes en la sede web de esta institución como otras muchas que la citan. Por el contrario si se realiza la misma búsqueda en el directorio Dmoz (The Open Directory Project: <http://dmoz.org/> ) se obtiene una relación de sólo 15 sitios web, es decir una selección de 15 enlaces que forman parte de una categoría considerada dentro del directorio: la sede web de esta institución y de cada una de sus publicaciones electrónicas en línea. Es decir, que a través del directorio no se accede a páginas dispersas, sino a sitios web formados por un conjunto de páginas que conforman en cierta medida una unidad documental.

Los buscadores ofrecen una mayor exhaustividad y actualización, pero en ocasiones pueden producir un efecto de dispersión por exceso de información. Por el contrario, las guías, índices o directorios de Internet pretenden ofrecer herramientas de precisión para la recuperación de información: intentan presentar información depurada, seleccionada y clasificada, para lo cual precisan de una mayor intervención de trabajo humano. Responden a necesidades de información diferenciadas: a través del buscador se pueden localizar datos dispersos, pero sólo a través de un directorio se puede obtener un listado organizado de puntos esenciales de referencia sobre un tema.

En los últimos años se ha producido un fenómeno creciente de portalización al cual no son ajenas las bibliotecas y centros de documentación. Estas instituciones no desean limitarse a mostrar sus propios recursos presentes en su sede web, sino que intentan actuar como intermediarios en el acceso a recursos externos. Algunos de ellos son gratuitos pero otros precisan de una suscripción para el acceso de los usuarios de la biblioteca, por lo que es necesario publicitar y crear instrumentos para orientar su uso.

Como afirman Abadal y Estivill [1] : "facilitar el acceso a un conjunto seleccionado de recursos web se puede considerar una tendencia plenamente consolidada en las bibliotecas de los países avanzados". Para ello pueden optar por tres opciones:

1. incluir en sus propios catálogos registros relativos a estos recursos, siguiendo las posibilidades que ofrecen los sistemas de gestión bibliotecaria.
2. crear guías de recursos de estructura sencilla y con una descripción elemental.
3. diseñar bases de datos más complejas específicamente diseñadas para la descripción de recursos seleccionados.

En este artículo se pretende analizar esta tercera opción, como un modelo concreto de directorio que ha sido desarrollado en bibliotecas, centros de documentación o consorcios de estas instituciones. Este modelo se denomina habitualmente como **Subject Gateways (pasarelas temáticas)** . Se trata de sistemas de información que seleccionan, describen y clasifican recursos web aplicando criterios propios de la Biblioteconomía y Documentación, como las clasificaciones disciplinares o los lenguajes controlados. Para Sebastià y Martínez-Carner [2] , las pasarelas temáticas se caracterizan por la aplicación del "control de calidad en la selección, descripción, indización y evaluación de los recursos". Son por tanto modelos de directorios dirigidos a la comunidad académica y científica, basados en la selección de recursos de información de alta calidad, identificados y evaluados por profesionales de la información (bibliotecarios o documentalistas especializados). Pueden presentarse dentro de una sede web propia, como servicio de un portal bibliotecario o formando parte de los contenidos de un portal científico.

La terminología aplicada a este tipo de bases de datos es muy variada. Además de Subject Gateways, reciben denominaciones diversas: Information Gateways, Subject Based Information Gateways (SBIG), Quality-controlled Subject Gateways, Internet Subject Directories o Resource Discovery Systems. En castellano pueden aparecer también como: bases de datos de recursos web [3] o directorios analíticos. A pesar de esta indefinición, su presencia en Internet está plenamente consolidada.

La separación entre pasarelas temáticas y otros sistemas de información sobre la web (directorios, bibliotecas virtuales de enlaces y guías temáticas de Internet) es a menudo confusa, y frecuentemente se encuentran agrupadas en enumeraciones como la recopilación de Pinakes ( <http://www.hw.ac.uk/libWWW/irn/pinakes/pinakes.html> ) . En todos estos modelos se aplica la clasificación de recursos por materias, elemento esencial cuya importancia ha sido analizada en el documento coordinado por Koch y May para el proyecto Desire [4] . Sin embargo, este trabajo se centra en las pasarelas temáticas que además de la clasificación, aportan una descripción más elaborada de recursos web.

## 2. Pasarelas temáticas y aplicación de las técnicas de análisis documental

Las pasarelas temáticas pueden considerarse nuevas aplicaciones de las técnicas de análisis documental desarrolladas para la construcción de bases de datos bibliográficas, las cuáles han sido de vital importancia para la comunidad científica en el último cuarto del siglo XX. El desarrollo de la edición electrónica en Internet hace suponer que estos recursos puedan quedar obsoletos, ya que a menudo ofrecen tan sólo referencias y no permiten interrogar el texto completo de los documentos. Sin embargo constituyen una importante tradición que ha desarrollado una técnicas documentales que seguirán siendo útiles en los futuros sistemas de información y que constituyen el modelo en el que se basan las pasarelas temáticas en Internet.

Según María Pinto Molina [5] el análisis documental se define como el "Conjunto de operaciones - intelectuales y mecánicas - que afectan al contenido y a la forma de los documentos originales, reelaborándolos y transformándolos en otros de carácter instrumental o secundario, que faciliten al usuario la identificación precisa, la recuperación y la difusión de aquellos". Define por tanto una metodología de trabajo utilizada para construir sistemas de información, que inicialmente fueron bibliografías y repertorios bibliográficos impresos y que posteriormente se transformaron en bases de datos bibliográficas.

Como conjunto de operaciones el análisis documental implica atender a la forma y el contenido de los documentos para constituir un registro que incorpore diferentes elementos:

1. **Descripción bibliográfica:** anotación normalizada de los elementos indispensables para constituir una referencia bibliográfica que permita identificar y localizar con seguridad un documento: título, datos de publicación, formato, autoría o responsabilidad, idioma, país de edición, tipo de documento...
2. **Resumen:** información esencial sobre la temática tratada por los documentos analizados, facilitando su conocimiento a través de una lectura rápida.
3. **Clasificación:** asignación de una notación extraída de un sistema predeterminado de organización del conocimiento estructurado en clases y subclases. Usualmente se utilizan sistemas de clasificación enciclopédicos universales en bases de datos multidisciplinares y sistemas especializados en bases de datos circunscritas a un ámbito disciplinar concreto.
4. **Indización:** extracción de un conjunto de conceptos que especifiquen los temas tratados que constituyen entradas en un índice de materias. Puede tratarse de términos de libre asignación (palabras clave) o pertenecientes a un tesauro o lenguaje controlado (descriptores o encabezamientos de materia normalizados). Frecuentemente las bases de datos optan por incluir varios campos para la indización, distinguiendo entre descriptores (conceptos) e identificadores (nombres propios de personas, organismos, obras) o incluso ampliando con otros campos para reflejar necesidades específicas de información (compuestos químicos, topónimos, periodos históricos, legislación).
5. Informaciones adicionales que enriquecen el valor de las bases de datos: afiliación institucional o lugar de trabajo de los autores, información sobre la disponibilidad de consulta de los documentos, etc.

Al igual que las bases de datos bibliográficas, la creación de una pasarela temática prescinde de la estructura de campos de las normas de catalogación (que toma en consideración solamente como modelo de referencia para la descripción) y de la utilización de un sistema de gestión bibliotecaria. Y así como hay grandes diferencias en la construcción de diferentes bases de datos bibliográficas, también encontraremos diferentes modelos de pasarelas temáticas. Los consorcios o instituciones creadores han de tomar sus propias decisiones en cuanto a:

- Definición de criterios de selección: qué se va a seleccionar, según tipo de recurso, su ámbito de audiencia, la procedencia geográfica, criterios de calidad, etc.
- Diseño de una base de datos con una adecuada estructura de campos, que incluyen diferentes elementos de análisis formal y de contenido:
  - Campos para la identificación y localización del recurso: título, autoría o responsabilidad, localización web (URL u otros), tipología de recurso, idioma,...
  - Clasificación jerárquica por disciplinas.
  - Indización por palabras clave o por descriptores extraídos de un tesauro (con otros tipos de índices adicionales en ocasiones).
  - Descripción del contenido (resumen).
  - Campos de control interno: analista, fecha de creación del registro y de la última consulta o actualización de datos,...
- Diseño de una interfaz de usuario: visualización de registros, pantallas de búsqueda simple y avanzada, así como la navegación a través del sistema de clasificación (browsing).

En consecuencia, no hay uniformidad entre las diferentes pasarelas temáticas, que pueden diferir en sus criterios de selección, su estructura de campos o su interfaz de consulta. Pero sí hay coincidencia en sus objetivos: selección de recursos de calidad y análisis de contenido basado en técnicas documentales.

### 3. Normalización y sistematización

#### 3.1. Normalización interna

Para que un sistema de información pueda cumplir con sus objetivos debe proceder de forma sistemática en la toma de datos. La normalización del proceso de análisis documental tiene por objetivo garantizar la consistencia del conjunto. Los productores deben esforzarse en documentar la toma de decisiones, redactando guías o manuales de normas para la cumplimentación de campos de la base de datos o para definir los criterios de calidad en la selección de recursos.

Para la sistematización de la toma de datos en campos específicos pueden crearse modelos propios o bien adecuarse a normas internacionales. Las normas ISBD pueden aplicarse a la forma de los nombres de autores o instituciones. Igualmente, pueden aplicarse los códigos internacionales para las lenguas (ISO 639), los nombres de países (ISO 3166) o para la representación de fechas (ISO 8601).

Los lenguajes documentales (Sistemas de Clasificación, Léxicos y Tesauros) son también herramientas para la normalización en el análisis del contenido. Las pasarelas temáticas suelen recurrir al uso de clasificaciones académicas. En sistemas multidisciplinares se pueden utilizar los sistemas de clasificación nacionales o las clasificaciones enciclopédicas de mayor difusión: Clasificación de la Library of Congress (LCC), Clasificación Decimal de Dewey (DDC) o la Clasificación Decimal Universal (CDU o UDC). En bases de datos especializadas pueden usarse clasificaciones específicas como por ejemplo la National Library of Medicine (NLM) en Biomedicina, o Engineering Information (Ei) en Ingeniería. El uso de clasificaciones internacionales facilita también la construcción de interfaces multilingües para la consulta.

En cuanto a los sistemas de indización algunas bases de datos han recurrido a la libertad de las palabras clave, mientras que otras han optado por la normalización de los lenguajes controlados como la Library of Congress Subject Headings (LCSH) para sistemas multidisciplinares, el Medical Subject Headings (MeSH, <http://www.nlm.nih.gov/mesh/>) en Biomedicina o el HASSET Thesaurus (<http://www.data-archive.ac.uk/search/hassetSearch.asp>) en Ciencias Sociales y Humanas. También puede recurrirse a tesauros solamente para campos específicos, por ejemplo en topónimos con el Getty Thesaurus of Geographical Names (<http://www.getty.edu/research/tools/vocabulary/tgn/index.html>).

En determinados ámbitos se carece de normas internacionales o lenguajes documentales especializados, o bien los que hay no resultan adecuados para responder a necesidades específicas de información. La creación de tablas, clasificaciones o léxicos de indización hechos a medida también implica una normalización que persigue la consistencia interna de una base de datos.

### 3.2. Normalización entre sistemas

La construcción de un sistema de información es un proceso costoso que se puede ver enriquecido por la cooperación interinstitucional. Si se desea contar con la posibilidad de compartir datos, importar y exportar registros, es necesario buscar la interoperabilidad. Para ello, serán necesarios dos condiciones:

1. Coincidencia en los criterios y herramientas utilizados para campos específicos: lenguajes documentales o códigos de idioma, país, fecha. Por ejemplo, el uso de clasificaciones universales o de varias clasificaciones, favorece la interoperabilidad con otros sistemas de información.
2. Compatibilidad en el diseño de la estructura de campos de acuerdo a normas internacionales. Para poder intercambiar datos, los diferentes sistemas deben coincidir al menos en la estructura básica que identifique los elementos esenciales de información.

La interoperabilidad estaría garantizada a través del formato MARC, para las instituciones que han optado por incluir la descripción de recursos web dentro de los catálogos. Sin embargo, las pasarelas temáticas se han diseñado al margen de este formato bibliográfico. Por tanto, para buscar la interoperabilidad se han de tener en cuenta los formatos relativos al uso de metadatos, específicamente diseñados para la documentación electrónica.

El movimiento DCMI (Dublin Core Metadata Initiative) especifica un conjunto estándar de metadatos para la descripción de recursos web, que ha sido aceptado como norma internacional ISO 15836-2003. Los elementos que componen la norma Dublin Core [6] son los siguientes:

- Título (DC.Title): nombre dado al recurso por el autor
- Autor o creador (DC.Creator)
- Materias (DC.Subject): clasificación o descriptores, indicando la fuente utilizada.
- Descripción (DC.Description): texto descriptivo o resumen
- Editor (DC.Publisher): entidad responsable de la edición web
- Colaborador (DC.Contributor): ilustrador, traductor, etc.
- Fecha (DC.Date): fechas de edición y sucesivas modificaciones
- Tipo de recurso (DC.Type)
- Formato (DC.Format)
- Identificador (DC.Identifier): URL u otros códigos que identifiquen el recurso.
- Fuente (DC.Source): procedencia
- Idioma (DC.Language)
- Relación (DC.Relation): otros recursos de los cuáles es parte, versión, etc.
- Cobertura (DC.Coverage): ámbito espacial o temporal

- Derechos (DC.Rights): especificación de los derechos de autor

Muchas pasarelas temáticas toman como modelo esta estructura de campos y la enriquecen con otros campos propios. De esta manera se garantiza la interoperabilidad, ya que se pueden exportar fácilmente sus registros utilizando las etiquetas normalizadas Dublin Core.

En otras pasarelas temáticas se han diseñado formatos propios de metadatos como el EELS Metadata Format de la Swedish Universities of Technology Libraries o el EEVL Format de la Edinburgh Engineering Virtual Library.

## 4. Ejemplos de pasarelas temáticas

En la web de la biblioteca de la Heriot Watt University puede consultarse una recopilación de directorios y pasarelas temáticas: <http://www.hw.ac.uk/libWWW/irn/pinakes/pinakes.html>.

A continuación se analizan por espacios geográficos, las iniciativas que aportan un mayor nivel de análisis de contenido.

### 4.1. Reino unido

Las bibliotecas especializadas del Reino Unido constituyen el mejor modelo de redes de cooperación que permite ofrecer a la sociedad trabajos conjuntos. El programa eLib, gestionado por el Joint Information Systems Committee (JISC), ha financiado y coordinado la creación de numerosas pasarelas temáticas. El proyecto ROADS (Resource Organisation and Discovery in Subject-based Services) ha desarrollado herramientas de software y guías metodológicas para la descripción de recursos.

Fruto de esta cooperación es la pasarela de la red **RDN (Resource Discovery Network)** : <http://www.rdn.ac.uk/> , un directorio multidisciplinar, que agrupa los registros suministrados por diferentes sistemas especializados que constituyen la red, a los cuáles se denomina como Hubs. Este término puede traducirse como "Concentradores" y hace referencia a la función de estos centros como recolectadores de información en sus áreas temáticas respectivas. Los Hubs proveen datos para el RDN, pero son también a su vez servicios independientes que ofrecen pasarelas a recursos de Internet en sus áreas temáticas y a menudo otros servicios añadidos. Un Hub puede ser una única institución o bien un consorcio de bibliotecas u organizaciones profesionales.

Los Hubs que se agrupan en la pasarela RDN son los siguientes:

- ALTIS <http://www.altis.ac.uk/> - Ocio, turismo y deporte
- Base coordinada por la University of Birmingham. El análisis de contenido incluye descripción, descriptores, clasificación por materias y por tipología de recursos.
- Artifact: <http://www.artifact.ac.uk/> - Bellas Artes, diseño
- Producto elaborado por un consorcio de bibliotecas universitarias. En sus registros incorpora descripción, palabras clave y clasificación por encabezamientos de materia (lista propia Artifact Subject Headings).
- Biome: <http://biome.ac.uk/> - Ciencias de la vida y la salud.
- Sistema de información coordinado por la Greenfield Medical Library de la University of Nottingham, que realiza descripción, palabras clave y clasificación por tipología de recursos. Este producto es a su vez fruto de la cooperación de otras pasarelas:
  - Omni: <http://omni.ac.uk/> - Medicina y salud. Usa la clasificación de la NLM.
  - VetGate: <http://vetgate.ac.uk/> - Veterinaria.
  - BioResearch: <http://bioresearch.ac.uk/> - Biomedicina y Biología.
  - Natural Selection: <http://nature.ac.uk/> - Mundo natural.
  - AgriFor: <http://agrifor.ac.uk/> - Agricultura, Forestales y Alimentación. Además, Biome aloja estas pasarelas:
  - Psci-com: <http://www.psci-com.org.uk/> - Política y comunicación científica.
  - MedHist: <http://medhist.ac.uk/> - Historia de la Medicina.
  - AHIS: <http://www.ahis.org/> - Salud animal
- EEVL: <http://www.eevl.ac.uk/> - Ingeniería, Matemáticas e Informática
- Realizada por un consorcio de bibliotecas coordinadas por la Heriot Watt University, esta pasarela incorpora en sus registros descripción, palabras clave, clasificación por materias y por tipología de recursos.

- GEsorce: <http://www.gesource.ac.uk/home.html> - Geografía y Medio Ambiente.
- Base gestionada por el consorcio de bibliotecas académicas de Manchester (CALIM). En sus registros incluye descripción, palabras clave, clasificación por materias y por tipo de recurso y localización geográfica (topónimos, latitud y longitud).
- Humbul Humanities Hub: <http://www.humbul.ac.uk/> -
- Pasarela desarrollada en la University of Oxford. En sus registros aporta descripción, clasificación por materias (Humbul Subjects) y por tipología de recursos, periodo histórico (según una tabla cerrada), tipo de destinatario (3 categorías) y descriptores de materias. Utiliza los metadatos Dublin Core como base para su estructura de campos.
- PSIGate: <http://www.psigate.ac.uk/newsite/> - Física
- Pasarela gestionada por el consorcio de bibliotecas académicas de Manchester (CALIM). Realiza descripción, palabras clave, clasificación por materias y por tipo de recurso.
- SOSIG: <http://www.sosig.ac.uk/> - Ciencias Sociales, Economía y Derecho Sus siglas responden a Social Science Information Gateway. Ha sido un sistema pionero desarrollado por el Institute for Learning and Research Technology de la University of Bristol. Sus registros incluyen descripción, palabras clave (extraídas de los tesauros HASSET, IBSS Y CareData), clasificación por tipología de recursos y por materias (basada en la CDU). SOSIG ofrece además una comunidad virtual especializada en ciencias sociales.

Además del conjunto de pasarelas temáticas del RDN, en el Reino Unido encontramos otras iniciativas de gran interés:

- ADAM: <http://adam.ac.uk/> - Bellas ArtesIniciativa del Surrey Institute of Art & Design y el Adam Consortium. Sus registros incluyen descripción, una clasificación propia (ASH, ADAM Subject Heading), una clasificación numérica (DDC), descriptores para conceptos generales (del Art & Architecture Thesaurus), periodo histórico (tabla cerrada), nombres de persona, instituciones, topónimos (Thesaurus of Geographic Names) y tipo de recurso (tabla cerrada).
- AERADE: <http://aerade.cranfield.ac.uk/> - Ingeniería aeroespacial y militarIniciativa de la Cranfield University que selecciona, evalúa y describe recursos, que pueden visualizarse en formato de directorio a través de una clasificación específica basada en la clasificación de la NASA y la del Military Science Index. A su vez actúa de proveedor de contenidos para la pasarela EELV de Ingeniería que forma parte de la red RDN. Su estructura y aplicación del software ROADS ha sido descrita por P. García Delgado. [7]
- AHDS (Arts and Humanities Data Service): <http://ahds.ac.uk/> - Humanidades y ArtePasarela coordinada por el King's College de Londres.
- Biz/ed: <http://catalogue.bized.ac.uk/> - Economía y negociosSistema de información creado por el Institute for Learning and Research Technology (ILRT), de la University of Bristol. Utiliza la clasificación Dewey (DDC), descripción y palabras clave.
- Bubl Link 5:15: <http://bubl.ac.uk/link/> - MultidisciplinarDirectorio gestionado por la Starthclyde University de Glasgow. En su análisis de contenido incluye descripción, clasificación (DDC), materias (basadas en LCSH), topónimos (países y regiones) y tipología de recursos (tabla cerrada). Para la clasificación comenzó utilizando la CDU, pero luego adoptó la DDC.
- Chemdex: <http://www.chemdex.org/> - Ciencias químicas.Pasarela creada en la University of Sheffield. Ofrece descripción, clasificación por materias y valoración de recursos.
- History on Line: <http://www.history.ac.uk/search/> - HistoriaIniciativa que colabora en los contenidos de Humbul.
- PORT: <http://www.port.nmm.ac.uk/> - Ciencias marítimasPasarela realizada por el National Maritime Museum de Greenwich. Recoge recursos sobre ciencias marítimas, desde un punto de vista pluridisciplinar, que describe y clasifica. Permite recuperar por palabras clave, periodos históricos y por tipo de recurso.

## 4.2. Resto de Europa

A nivel europeo se han desarrollado varias iniciativas cooperativas para desarrollar herramientas y consensuar criterios comunes para la creación de pasarelas temáticas. En 1998 se inició el proyecto DESIRE: Development of a European Service for Information on Research and Education ( <http://www.desire.org/> , <http://www.lub.lu.se/desire/> ) en el que han colaborado el Reino Unido, Países Bajos, Noruega y Suecia.

El principal resultado de la cooperación europea es actualmente el sistema de información Renardus ( <http://www.renardus.org/> ), creado por un consorcio europeo, en el que participan Alemania, Dinamarca, Finlandia, Países Bajos y Reino Unido. En este producto se pueden interrogar conjuntamente los registros que provienen de las pasarelas participantes:

- Anglistik Guide: Virtual Library of Anglo-American Culture - Literature at SUB Göttingen (GoeAng): <http://www.anglistikguide.de/>
- DAINet - German Agricultural information Network: <http://www.dainet.de/>
- Danmarks Elektroniske Forskningsbibliotek (deff): <http://deff.dk/>
- DutchESS - Dutch Electronic Subject Service: <http://www.kb.nl/dutchess/>
- Finnish Virtual Library (FVL): <http://www.jyu.fi/library/virtuaalikirjasto/>
- Geo-Guide: Virtual Library of Earth Sciences, Geography, Thematic Maps, and Mining at SUB Göttingen (GoeGeo): <http://www.geo-guide.de/>
- History Guide: Virtual Library of Anglo-American Culture - History at SUB Göttingen (GoeHist): <http://www.historyguide.de/>
- MathGuide: Part of the Virtual Library of Mathematics at SUB Göttingen (GoeMath): <http://www.mathguide.de/>
- NOVAGate - Nordic Gateway to Information in Forestry, Veterinary and Agricultural Sciences: <http://novagate.nova-university.org/>
- Online theses, Germany (DNB-Theses): <http://deposit.ddb.de/>
- Resource Discovery Network (RDN): <http://www.rdn.ac.uk/>

Otro proyecto cooperativo europeo es Lingu@net Europa (<http://www.linguanet-europa.org/y2/>) creado por un consorcio en el que participan, por parte española, el Instituto Cervantes y la Universidad Politécnica de Madrid. L. Bellido y E. Pastor [8] han descrito el uso de metadatos en este proyecto.

Entre las iniciativas nacionales cabe destacar varios productos en los países nórdicos: Biogate (<http://biogate.lub.lu.se/>), EELS (<http://eels.lub.lu.se/>), la biblioteca virtual finesa (<http://www.jyu.fi/library/virtuaalikirjasto/>) o NOVAGate (<http://novagate.nova-university.org/>).

En los Países Bajos es la Biblioteca Nacional la que desempeña una función de liderazgo en la cooperación con bibliotecas universitarias para el desarrollo de la pasarela DutchESS: Dutch Electronic Subject Service (<http://www.kb.nl/dutchess/index.html>) que participa en Renardus. En este sistema se incluye descripción, clasificación por materias (según el sistema nacional Dutch BC) y por tipología de recursos.

En Alemania en la página de la Biblioteca de Göttingen, dedicada a la Special Subject Guides SSG-Fachinformation (SSG-FI): <http://www.sub.uni-goettingen.de/ssgfi/>, se puede acceder a sus diferentes bases de datos: MathGuide, GeoGuide, Anglo-American Culture y ForestryGuide. Tomando como ejemplo la Virtual Library of Anglo-American Culture (<http://www.sub.uni-goettingen.de/vlib/>), se aporta descripción y materias según la LCSH. Además de estas bases, en Renardus participa Dainet (<http://www.dainet.de/>), especializada en Agricultura. Otra iniciativa alemana es el directorio GERHARD, German Harvest Automated Retrieval and Directory (<http://www.gerhard.de/>). Este sistema clasifica de forma automática los recursos alemanes con los códigos d! e la CDU.

En Francia también se puede citar como ejemplo CISMef (<http://www.cismef.org/> o <http://www.chu-rouen.fr/cismef/>) del Hospital Universitario de Rouen, que incluye descripción, materias (MeSH) y tipología de recursos (tabla cerrada).

### 4.3. Estados unidos

En Estados Unidos las bibliotecas universitarias han optado generalmente por la catalogación y por la elaboración de guías de recursos con una descripción muy somera. OCLC ha desarrollado sucesivos programas de catalogación de recursos web: NetFirst y CORC (Cooperative Online Resource Catalog), que en 2002 se integró dentro del servicio de catalogación cooperativa Connexion (<http://www.oclc.org/connexion/>). Este sistema de catalogación compartida permite exportar registros tanto en formato MARC como en metadatos Dublin Core.

Como ejemplos de pasarelas temáticas desarrolladas en Estados Unidos con mayor nivel de análisis del contenido se encuentran:

- AgNIC: <http://www.agnic.org/>
- Sistema creado por el Agriculture Network Information Center (National Agricultural Library), que incluye descripción, clasificación propia, descriptores (NAL Agricultural thesaurus) y palabras clave.
- CyberStacks: <http://www.public.iastate.edu/~CYBERSTACKS/>
- Base de datos multidisciplinar que recopila bases de datos y recursos de búsqueda que describe y clasifica según la LCC.

- Health Web: <http://healthweb.org/>
- Sistema creado por un consorcio de bibliotecas médicas que incluye descripción y clasificación.
- Infomine: <http://infomine.ucr.edu/>
- Recopilación de recursos académicos de la University of California, en colaboración con otras universidades americanas. En su análisis del contenido realiza descripción, clasificación por materias (LCC) y por tipología de recursos, descriptores de materias (LCSH) y palabras clave.
- Internet Library for Librarians: <http://www.itcompany.com/info retriever/>
- Pasarela gestionada por la empresa Infoworks Technology Company. Incluye clasificación por tipología de recursos, descripción y palabras clave.
- Librarians' Index to the Internet: <http://lii.org>
- Sistema creado por la Library of California, con descripción y materias (LCSH).
- Medical Matrix: <http://www.medmatrix.org/>
- Servicio comercial que incluye descripción, valoración, clasificación y materias (MeSH).

#### 4.4. Australia y nueva zelanda

Las pasarelas australianas son, al igual que las británicas, un modelo de trabajo cooperativo, bajo la coordinación de la National Library. En la sede web del grupo de trabajo de metadatos The Dublin Core Australia/New Zealand ( <http://dcanzorg.ozstaging.com/mb.aspx> ) se enumeran 26 pasarelas temáticas australianas y una de Nueva Zelanda. También se puede acceder a documentos como una guía de buenas prácticas para las pasarelas temáticas australianas: <http://www.nla.gov.au/initiatives/sg/bestpractice.html>

Un ejemplo de pasarela australiana es Agrigate ( <http://www.agrigate.edu.au/> ), un servicio especializado en ciencias agrarias, mantenido de forma cooperativa por las bibliotecas universitarias de Melbourne, Adelaide y Queensland y bibliotecas de investigación de SCIRO. Realiza fichas descriptivas de los recursos de acuerdo a la estructura de campos de Dublin Core, a la que se añaden algunos elementos propios que se enumeran en <http://www.agrigate.edu.au/metadate.html> . Para la indización de materias utilizan los tesauros CAB International y Agterms.

#### 4.5. España

En España la mayor parte de las bibliotecas universitarias ofrecen recursos web dentro de sus catálogos y numerosas guías de recursos web, con una somera descripción. Ejemplo de ello son las guías temáticas que participan en el catálogo colectivo de las universidades catalanas ( <http://ccuc.cbuc.es/vtls24/vtls/catalan/consultagt.html> ) o Complured ( <http://alfama.sim.ucm.es/rie/framerie.htm> ) de la Biblioteca de la Universidad Complutense de Madrid.

En 1999 siete universidades españolas iniciaron el proyecto cooperativo Darwin ( <http://www.kronosdoc.com/darwin/> ), Directorio Analítico de Recursos Web Informativos. Este proyecto [9] utilizó una clasificación adaptada de la CDU, materias, descripción y valoración de recursos.

En la Biblioteca de la Universidad Politécnica de Catalunya, el portal Bibliotècnica ofrece la consulta de una base de recursos web con descripción, clasificación por materias y por tipología de recursos. Por su parte, la Biblioteca General de la Universidad Politécnica de Valencia ha desarrollado la base de datos Rodamón ( <http://www.upv.es/bib/recursos/rodamon.htm> ), que incluye resumen, palabras clave y tipología de recursos. La estructura de campos se basa en el formato Dublin Core.

En Valencia se ha desarrollado también el proyecto QReME [10] : Biblioteca Médica Virtual ( <http://www.comv.es/> ), de acceso restringido a los colegiados del Colegio Oficial de Médicos de Valencia. Incluye descripción, clasificación, palabras clave y tipología de recursos.

En los grupos de trabajo de las comunidades virtuales de Rediris, se gestó el proyecto SARAC ( <http://rtr.rediris.es/cgi-bin/rtrnav?v=sarac> ) : Servicio de Acceso a Recursos de Alta Calidad. Se ha desarrollado con la colaboración de documentalistas de la Universidad de Granada e incluye descripción, palabras clave y clasificación por materias.

En el CINDOC se han elaborado diferentes bases de datos de recursos web como parte de los contenidos que se ofrecen en los portales científicos en desarrollo ( <http://www.cindoc.csic.es/servicios/portales.html> ). En la elaboración del directorio de recursos del portal Tecnociencia [11] ( <http://www.tecnociencia.es> ), coordinado por la FECYT y patrocinado por el MCYT, se ha incluido descripción, palabras clave, clasificación por tipología y clasificación por materias. Se han utilizado varios sistemas de clasificación para integrar registros de diferentes orígenes: ANEP,

## 5. Problemática de la construcción de pasarelas temáticas

El objetivo de las bibliotecas y centros de documentación es mantener una función de intermediación entre los usuarios y las fuentes de información, realizar una función de selección y orientación ante los problemas de credibilidad y usabilidad de la red. Para ello, muchas instituciones y consorcios han abordado proyectos de creación de sistemas de información que describen y analizan recursos web. Pero no todos han logrado consolidarse y mantenerse actualizados.

El modelo de pasarelas temáticas ofrece aún muchas dudas y puede quedar obsoleto ante la aparición de nuevos sistemas de recuperación de información. Sin embargo, las subject gateways han contribuido a la capacidad actual de búsqueda en Internet, facilitando la tarea de los motores de búsqueda. Desde el punto de vista de las técnicas documentales han aportado una nueva utilidad para los sistemas de indización y clasificación, como recurso para la mejora de la usabilidad en la navegación. Especialmente han servido para realzar la función y vigencia de los tradicionales sistemas de clasificación, a menudo infrautilizados en las bases de datos bibliográficas y los OPAC de bibliotecas.

El diseño de una pasarela temática plantea problemas concretos que se analizan a continuación:

### 5.1. Falta de acuerdo en cuanto a los campos de las bases de datos

Persisten grandes diferencias en el diseño de campos entre diferentes pasarelas temáticas. En el análisis formal, los únicos campos coincidentes son Título y URL (o URN). En ADAM (<http://adam.ac.uk/>) éste último se denomina URI para abarcar cualquier código de identificación de un recurso (URL, ISSN, etc.). En otros sistemas se incluye Autoría o Responsabilidad. Bubl Link (<http://bubl.ac.uk/link/index.html>) incorpora Tipo de recurso y Localización (país). En Humbul (<http://www.humbul.ac.uk/>) se añade Título alternativo, Fecha de creación y modificación, Idioma, Audiencia, Edición y Otros datos (es citado por, es parte de, requerimientos técnicos).

En el análisis de contenido las pasarelas temáticas se diferencian de las simples guías por su descripción más extensa y por utilizar una clasificación jerárquica por materias, en lugar de una tabla simple de materias. Algunos sistemas enriquecen el análisis de contenido con palabras clave o con descriptores extraídos de un tesoro, pero otros añaden campos específicos. Por ejemplo, Humbul incluye Cobertura geográfica y Cobertura por periodo. ADAM distingue entre la clasificación (DDC), las materias del Adam Subject Headings, Periodo histórico, Nombres de persona, Instituciones, Topónimos y Conceptos generales. GEsorce (<http://www.gesource.ac.uk/>) incorpora a la localización geográfica campos específicos para la latitud y la longitud.

Finalmente sólo algunos sistemas permiten visualizar los campos de control interno del registro, como la identificación del analista y la fecha de chequeo o de última modificación.

### 5.2. Dificultad en la definición de la unidad documental y la tipología de recursos

Las bases de datos bibliográficas suelen incluir un campo para distinguir tipos de documentos y poder limitar las búsquedas entre monografías, artículos, reseñas, documentos de trabajo o tesis. Cada registro de una base de datos se corresponde con una unidad documental, con cada artículo de una revista o con cada ponencia de un Congreso. Con ello el usuario puede conocer y seleccionar qué tipo de documentos quiere consultar. Esta misma dinámica se ha intentado traspasar a las pasarelas temáticas, pero se haya dificultada por la excesiva amplitud de la tipología de recursos y la dificultad para delimitarlos. No hay relación entre unidad documental y página web. Dentro de una sede web se pueden encontrar diferentes unidades o conjuntos de información: departamentos, bibliotecas, bases de datos en línea, publicaciones electrónicas. Los buscadores sólo distinguen páginas, cualquier guía de enlaces, directorio o pasarela temática desmenuza, selecciona y agrupa los elementos, según una tabla de tipología predeterminada.

La norma Dublin Core (<http://es.dublincore.org/documents/dcmi-type-vocabulary/>) sólo especifica una tipología muy genérica de contenidos, con estos valores: Collection, Dataset, Event, Image, Interactive Resource, Service, Software, Sound, Text, Physical Object, Still Image y Moving Image.

La pasarela Bubl Link (<http://bubl.ac.uk/link/index.html>) distingue una tipología con veintinueve categorías:

Bibliographies Biographies Bookshops Dictionaries Mailing lists Reference data, reference documents Documents Thesauri	Book and Text Collections Essays Guides, Tutorials, Teaching Materials Journal Collections Journal Listings Magazine Listings Poems Proceedings
JISC Funded Services and Projects Library Catalogues UK University Home Pages UK Higher Education Colleges Other UK Institutions Government Worldwide	Image Collections Map Collections Moving Images Museums Satellite Images Software Archives Sound Collections

Un número similar de categorías pueden encontrarse en las tipologías de la pasarela Humbul (<http://www.humbul.ac.uk/about/catalogue.html>). Por el contrario un sistema como Infomine ha optado por un número muy reducido de opciones, agrupando categorías:

- Ejournals, periodicals, serials, news, letters
- Database, data set, time series, image-base
- Directories
- Virtual libraries, search engines
- Reference resources
- Instructional uses, textbooks manuals, classes, courses, curriculum, study teaching
- Books, etext, text archives
- Maps & mapping

### 5.3. Necesidad de fijar unos criterios de selección

Frente a la exhaustividad de los buscadores, las pasarelas temáticas apuestan por la pertinencia de los contenidos. Se hace por tanto indispensable fijar unos criterios de evaluación de la calidad para delimitar una pautas de discriminación en la selección. Como señala L. Codina [12], los indicadores de calidad deben tener en cuenta sobre todo aspectos relativos al contenido y a la autoría explícita. Sin embargo, no resulta fácil establecer parámetros de análisis sin cierta ambigüedad ni mantener un sistema de trabajo que garantice evaluaciones periódicas de los recursos.

El proyecto DESIRE (<http://www.desire.org/results/discovery/>) publicó en 1997 un análisis de los criterios utilizados [13] y elaboró una encuesta de autoevaluación basada en 193 parámetros. Las conclusiones y propuestas de este programa han servido de modelo para perfilar la política de calidad de las pasarelas británicas. En Internet se pueden localizar algunos ejemplos de guías para la evaluación y selección de sitios web:

- ALTIS: <http://altis.ac.uk/guidelines/evaluation/>
- Biome: <http://biome.ac.uk/guidelines/eval/>
- Darwin: <http://exlibris.usal.es/merlo/darwin/material/guia.htm>
- DutchESS: [http://www.kb.nl/dutchess/manual/quality\\_eng.html](http://www.kb.nl/dutchess/manual/quality_eng.html)
- EELS: <http://www.lub.lu.se/eel/qualcrit.html>
- GEsorce: <http://www.gesource.ac.uk/content3v4.doc>

### 5.4. Preocupación por la usabilidad y recuperabilidad

La eficacia en la recuperación de información no se puede garantizar en sistemas probabilísticos en los que no se puede predecir o limitar las preguntas que se realicen al sistema. Sin embargo, las técnicas documentales deben ayudar a mejorar las opciones de recuperación, ya que permiten diseñar una interfaz de búsqueda avanzada, con opciones de búsqueda limitada por campos como institución, descripción, palabras clave, etc. La clasificación por materias y por tipología de recurso permiten la navegación en formato de directorio jerárquico y la limitación de las búsquedas (selección de filtros para realizar la búsqueda por palabras dentro de un bloque o epígrafe de una tabla predeterminada).

El resumen o descripción permite dar un valor añadido a la visualización de los registros para orientar mejor al usuario.

Para mejorar la usabilidad debe ser muy sintético, rico en contenidos precisos pero a la vez muy breve. La indización por descriptores además de ser una opción de búsqueda es también un medio para orientar la ponderación en la salida de los resultados. La pertinencia de un concepto presente en el campo de descriptores o palabras clave tiene una mayor fiabilidad que la presencia del mismo en el contenido de otros campos como la descripción (donde en ocasiones se citan aspectos secundarios) o la clasificación (donde en ocasiones se agrupan varios aspectos en la redacción de un único epígrafe clasificatorio).

Evidentemente, un directorio no puede competir con la exhaustividad de un motor de búsqueda aplicado sobre el texto completo de las páginas web. Las pasarelas temáticas son bases de datos referenciales. Una opción de futuro puede ser la confluencia con los motores de búsqueda, de forma que pueda aprovecharse su potencia en la recuperación con la pertinencia de los lenguajes documentales, tesauros y clasificaciones. La cooperación interinstitucional y las redes de cooperación también pueden aportar nuevas soluciones de tratamiento automático y pertinencia de la recuperación limitada a un entorno en el que se garantice la veracidad de los metadatos incorporados en las propias páginas web y capturados sistemáticamente por un programa apropiado (harvester).

## 5.5. Inseguridad del mantenimiento

Actualmente la construcción de pasarelas temáticas precisa de una importante inversión de recursos humanos. Se necesita una alta dedicación, tanto durante el proceso técnico del análisis como para la actualización de contenidos: detección de nuevos recursos y modificaciones en los registros ya seleccionados. Internet es un medio demasiado inestable para garantizar la rentabilidad de esta inversión. Cabe preguntarse si las subject gateways constituyen un modelo obsoleto para la recuperación de información.

Sin embargo, el factor humano debe ser valorado como un valor decisivo e irrenunciable cuando se trata de apostar por una selección de recursos basada en criterios de calidad. Los productores de pasarelas temáticas deben plantearse la amortización de esta inversión sin renunciar a la aportación de los bibliotecarios y documentalistas. Para ello se puede trabajar en tres direcciones complementarias:

1. buscando la especialización, limitando la cobertura de las pasarelas temáticas a objetivos razonables, útiles y controlables;
2. desarrollando nuevas herramientas informáticas que faciliten los procesos técnicos;
3. aunando fuerzas con la creación de consorcios institucionales y comunidades de usuarios que contribuyan al mantenimiento y actualización de contenidos.

Es por tanto el modelo británico el mejor ejemplo a seguir: apoyo institucional (JISC) y tecnológico (ROADS) para grupos de trabajo especializados por disciplinas (Hubs), que forman parte de un consorcio nacional (RDN) y que a su vez participa en un sistema de información europeo (Renardus). En el Reino Unido se ha conseguido un sistema sostenible que garantiza la continuidad de un modelo que debe constituir una de las bases del desarrollo futuro de nuevas formas de recuperación de información en Internet.

**Nota:** Todas las direcciones de Internet incluidas en el texto del artículo han sido chequeadas el 30-01-2004

## 6. Notas

[1] Abadal, Ernest; Estivill, Assumpció. L'accés als recursos web des de les biblioteques. Informe encarregat pel Consorci de Biblioteques Universitàries de Catalunya. Barcelona: 2000 <http://www.cbuc.es/5digital/ireweb.pdf> [accesible el 26-01-2004]

[2] Sebastià Salat, Montserrat; Martínez-Carner Ascaso, Pilar. Criterios de Evaluación de la Interacción Indización/Recuperación de la Información en las Pasarelas Temáticas. En: JOTRI 2003: II Jornadas de Tratamiento y Recuperación de la Información, Leganés, 2003 . Madrid: Universidad Carlos III de Madrid, Escuela Politécnica Superior, Departamento de Biblioteconomía y Documentación y Departamento de Informática, 2003, pp. 239-249.

[3] Estivill, Assumpció; Abadal, Ernest. Acceso a los recursos web gratuitos desde las bibliotecas. El Profesional de la Información , 2000, 9 (11): pp. 4-20

[4] Koch, Traugott; Day, Michael (et al.) DESIRE Specification for resource description methods, Part 3: The role of classification schemes in Internet resource description and discovery. DESIRE (s.f.) [http://www.lub.lu.se/desire/radar/reports/D3.2.3/class\\_v10.html](http://www.lub.lu.se/desire/radar/reports/D3.2.3/class_v10.html) [accesible el 26-01-2004]

[5] María Pinto Molina: Análisis documental: fundamentos y procedimientos . Madrid: Eudema, 1991.

[6] DCMI: Conjunto de elementos Dublin Core, versión 1.1: Descripción de referencia. <http://es.dublincore.org/documents/dces/> [accesible el 27-01-2004]

- [7] García Delgado, Purificación. ROADS. Software y conjunto de normas para la gestión de recursos electrónicos en Internet. La experiencia en un caso concreto : AERADE portal temático aeroespacial . II Jornadas Españolas de Bibliotecas Digitales, 2001.
- [8] Bellido, Luis; Pastor, Encarna. Lingu@net Europa: Un catálogo de recursos multilingüe con garantía de calidad. En: I Congreso online del Observatorio para la CiberSociedad, 2002.  
<http://cibersociedad.rediris.es/congreso/comms/c02bellido.htm> [accesible el 28-01-2004]
- [9] Merlo Vega, José Antonio; Gracia Armendáriz, Juan; Zapico Alonso, Felipe F.; Rodríguez Gairiu, Josep Manuel. Darwin: Una propuesta de organización y evaluación del conocimiento accesible en línea. En: IV Congreso ISKO-ESPAÑA: La representación y la organización del conocimiento en sus distintas perspectivas . Granada: Capítulo Español de la Sociedad Internacional para la Organización del Conocimiento, 1999, pp. 87-94.
- [10] Abad García, M. Francisca; González Teruel, Aurora. Creación, mantenimiento y perspectivas de una base de datos de recursos web médicos: La experiencia de QReME. En: JOTRI 2003: II Jornadas de Tratamiento y Recuperación de la Información, Leganés, 2003 . Madrid: Universidad Carlos III de Madrid, Escuela Politécnica Superior, Departamento de Biblioteconomía y Documentación y Departamento de Informática, 2003, pp. 41-49.
- [11] Equipo de Trabajo Portaltecnociencia. Metodología para la creación de un sistema de información científico-técnica en Internet: el caso de PortalTecnoCiencia. En: Los sistemas de información en las organizaciones: eficacia y trasparencia. VIII Jornadas españolas de Documentación. Barcelona: Col·legi Oficial de Bibliotecaris i Documentalistes de Catalunya - FESABID, 2003, pp. 343-355.
- [12] Codina, Lluís. Evaluación de recursos digitales en línea: conceptos, indicadores y métodos. Revista Española de Documentación Científica , 2000, 23 (1), pp. 9-44. [http://bdoc.csic.es:8080/basisbwdocs\\_rdisoc/rev0001/2000\\_vol23-1/2000\\_vol23-1\\_pp9-44.htm](http://bdoc.csic.es:8080/basisbwdocs_rdisoc/rev0001/2000_vol23-1/2000_vol23-1_pp9-44.htm) [accesible el 30-01-2004]
- [13] Hofman, Paul; Worsfold, Emma (et al.) . Selection Criteria for Quality Controlled Information Gateways. DESIRE, 1997. <http://www.ukoln.ac.uk/metadata/desire/quality/> [accesible el 28-01-2004]