

PROCESOS DE DIGITALIZACIÓN Y DE RECUPERACIÓN
DE MATERIAL EN GALICIA

Javier Agenjo Bullón
Empresa DIGIBIS

[En este artículo, revisado a principios de febrero de 2006, actualizo algunos de los datos y referencias de determinadas informaciones que presenté y expuse el 1 de diciembre de 2005 porque me parece útil recoger estas actualizaciones en aspectos que podrían considerarse básicos].

Los procesos de digitalización y recuperación de material bibliográfico en Galicia seguramente tienen su antecedente en la Ley de Patrimonio Histórico Español y en el instrumento de control, que, definido en su Reglamento, dio origen al Catálogo Colectivo del Patrimonio Bibliográfico, proyecto que se lleva a cabo por medio de la colaboración entre las distintas Comunidades Autónomas y la Administración Central. La Comunidad Autónoma de Galicia fue una de las primeras en comenzar esta tarea, en 1989. Se han publicado a lo largo de estos años algunas revisiones del estado de la cuestión del proyecto¹, que permiten conocer su función y planteamiento. Un planteamiento estratégico sobre cuál ha de ser, desde mi punto de vista, la función de las Biblio-

1. «El catálogo colectivo del Patrimonio Bibliográfico Gallego: impresos de los siglos XVII y XIX» / Ignacio Cabano Vázquez, Xavier Agenjo Bullón, Pilar Palá Gasós, *Boletín de la ANABAD*, ISSN 0210-4164, tomo 38, n.º 1-2, 1988, pp. 249-258 y «El catálogo colectivo del Patrimonio Bibliográfico de Galicia (C.C.P.B.G.): Balance de una década en Galicia (1987-1997) y propuestas para el futuro» / Ignacio Cabano Vázquez, Helena Verdía Armada, Xavier Agenjo Bullón, *Boletín de la ANABAD*, ISSN 0210-4164, tomo 50, n.º 2, 2000, pp. 101-118.

tecas Nacionales a la hora de reunir digitalmente la información disponible tuvo ocasión de expresarla en una ponencia² realizada en las *Jornadas sobre Bibliotecas Nacionales*, que tuvo lugar en la Biblioteca Valenciana y que, por ser muy fácilmente accesible en la Web, me permito remitir al lector.

Sin duda alguna, otra de las actuaciones fundamentales para llevar a cabo la digitalización y la difusión del patrimonio bibliográfico y, en general, de todo aquello que constituye el legado que se recoge en las distintas Instituciones de Memoria, archivos, bibliotecas, museos, así como centros de documentación y hemerotecas está tomando una nueva fisonomía mediante la irrupción en la red de un nuevo tipo de bases de datos, en el cual la información se transforma en conocimiento y que reciben el nombre de ontologías. Si cuando pronuncié la conferencia el 1 de diciembre de 2005 la norma ISO 21127³ se encontraba en fase de borrador final, ya ha adquirido su forma definitiva y, sin duda, debe constituir el marco de referencia final sobre el que construir el edificio de la digitalización de los materiales bibliográficos, archivísticos y museológicos que propicien el acceso a la información y al documento y que permitan llevar a cabo de forma coordinada este tipo de iniciativas.

Me adelanto a razonamientos posteriores al afirmar que no sólo por la enorme riqueza de información que proporciona la prensa histórica, sino por la extremada fragilidad del soporte en el que se encuentra impresa y el deterioro espontáneo que sufre a causa de su pH ácido, constituye la primera prioridad en la recuperación de la información, sin contar como digo, en la enorme fuente de información que supone para un conjunto de amplia gama de intereses por parte no sólo de investigadores y especia-

2. «Recursos digitales: un reto para las bibliotecas nacionales» // *Jornadas sobre Bibliotecas Nacionales «Las Bibliotecas Nacionales del siglo XXI»*, Biblioteca Valenciana, 18-21 de mayo de 2005. - bv.gva.es/documentos/Ponencias/Agenjo.pdf.

3. Sobre la aprobación y sus repercusiones de la ISO/FDIS 21127 (que era el estado en que se encontraba cuando presenté mi exposición el 1 de diciembre de 2005, puede verse *Report on ISO 21127, Information and documentation – A reference ontology for the interchange of cultural heritage information: Plenary meeting of TC46 SC4 WG9, Chiang Mai, Thailand, February 8, 2006* / Nicholas Crofts, Convenor. – <http://www.niso.org/international/SC4/N581.pdf>.

listas, sino del público en general⁴. Por ello, siento una especial satisfacción al comprobar que, como ha expuesto en estas mismas *Xornadas* mi querido amigo y compañero don Daniel Buján, jefe del Servicio del Libro y Bibliotecas de la «Consellaría de Cultura de la Xunta de Galicia», la prensa gallega se ha beneficiado de una atención preferente y puede consultarse en el sitio que se señala a continuación. También la Biblioteca Virtual de Prensa Histórica, proyecto en curso de realización por la Subdirección General de Coordinación Bibliotecaria, ha prestado una atención coordinada con la Xunta de Galicia a la digitalización de la prensa gallega.



<http://www.csbg.org/bibliotecadixital/asp/index.asp>

Junto a las publicaciones periódicas ha sido siempre un objeto especial de atención el fondo antiguo, fundamentalmente las monografías, aunque también otros materiales, como la cartografía, se han visto favorecidos por la digitalización que se ha llevado a cabo en Galicia. No deja de encontrarse aquí una para-

4. «La cuarta salida de El Monje digital y sus problemas hemerográficos: una recapitulación» / Xavier Agenjo Bullón, *Boletín de la ANABAD*, ISSN 0210-4164, tomo 54, n.º 4, 2004, pp. 119-138.

doja, puesto que el fondo antiguo impreso, en líneas generales, en papel de buena calidad corre menos peligro de desaparecer que la prensa o incluso las monografías impresas desde 1850 aproximadamente; sin embargo, la rareza de estos fondos hace a veces aconsejable su digitalización no tanto para facilitar su conservación como su acceso. Por estos dos conceptos, preservación y accesibilidad, hasta hace poco tiempo antitéticos, se han visto felizmente resueltos en la síntesis que supone la digitalización, siempre y cuando ésta se lleve a cabo cumpliendo determinada normativa, extendida internacionalmente. Fundamentalmente, y como veremos más adelante, se trata de incorporar al modelo de digitalización la utilización de metadatos para, mediante el modelo OAI-PMH⁵, asegurar su visibilidad en la red por medio de métodos de recolección (*harvesting*) de metadatos y, en general, adoptando la normativa que, a instancias de la IFLA, de la ICA y del ICOM, suscribió la UNESCO basándose en las recomendaciones de la *Digital Library Federation*⁶.

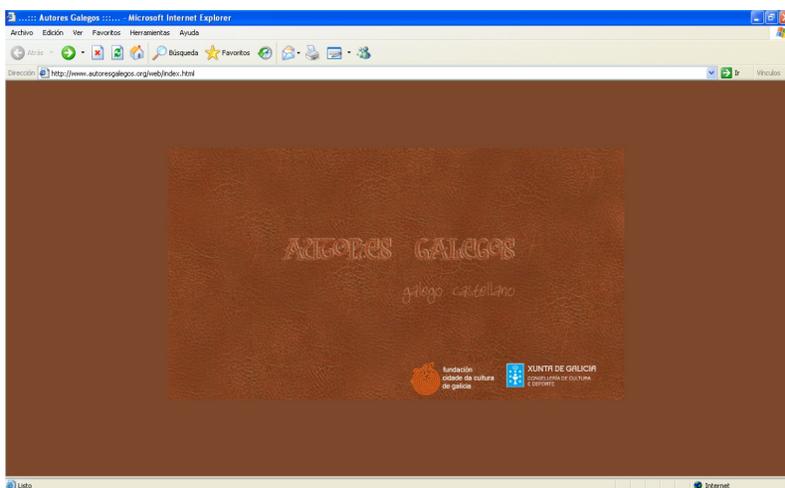
Una iniciativa importante, también ya realizada en Galicia, ha sido la elaboración de una bibliografía de autores gallegos, e, incluso ya de una forma continuada, de una base de datos en soporte cederrón, que recoge la producción gallega hasta 2002⁷.

5. «La recolección de metadatos (metadata harvesting) y su aplicación en España» / Xavier Agenjo Bullón y Francisca Hernández Carrascal // *9 Jornadas Españolas de Documentación*, Madrid, 14 y 15 de abril de 2005, Palacio Municipal de Congresos; [organiza, FESABID; coordina, SEDIC]. – [Madrid]: FESABID, 2005. – 586 p. – ISBN 84-930335-5-3. Además del texto de la comunicación, puede verse la presentación, que tuvo lugar el viernes, 19 de abril, http://www.fesabid.org/madrid2005/descargas/presentaciones/comunicaciones/hernandez_francisca.pps.

6. *Directrices para proyectos de digitalización: de colecciones y fondos de dominio público en particular para aquellos custodiados en bibliotecas y archivos* / estas directrices son el resultado del trabajo de un grupo de expertos de IFLA e ICA (International Council on Archives) invitados por la UNESCO; traducidas por el grupo de trabajo de colecciones digitales de las Comunidades Autónomas y el Ministerio de Cultura. – Madrid: Ministerio de Cultura, Secretaría General Técnica, D.L. 2005. – 183 p.; 21 cm – ISBN 84-8181-271-4.

7. *Centro Superior Bibliográfico de Galicia. O libro galego (ata 2003)*. [Recurso electrónico] / Elabora, Centro Superior Bibliográfico de Galicia. – [Santiago de Compostela]: Xunta de Galicia, Dirección Xeral de Promoción Cultural, [2005]. – 1 CD-ROM. – (*Patrimonio bibliográfico de Galicia. Serie bibliográfica*; 1). – ISBN 84-453-4033-6.

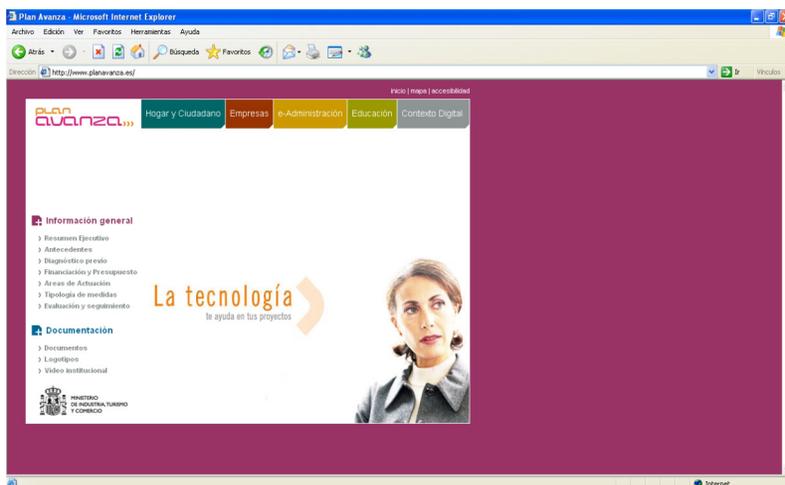
Es muy probable que esta información se actualice próximamente, de tal modo que se pueda disponer de una bibliografía nacional retrospectiva. Es importante señalar que todos estos materiales digitalizados, además de su difusión a través de la red, bien acudiendo directamente a las URL en las que se encuentran o bien mediante procedimientos más avanzados como repositorios OAI, o incluso ontologías, pueden constituir perfectamente un conjunto de materiales para la elaboración de recursos didácticos, de tal modo que el discente pueda acceder desde dichos *recursos didácticos* a los propios libros, a los documentos de archivo o a las piezas de museo, teniendo así una imagen más cabal y perfecta de su patrimonio propio, lo que facilitará la adquisición de su conocimiento.



<http://www.autoresgalegos.org/>

Una de las mayores dificultades que presenta la edición digital en la Web es pasar del estado de las ediciones facsimilares, en modo imagen, a las ediciones en modo texto, pues el reconocimiento óptico de caracteres no funciona tan eficazmente como quisiéramos cuando se trabaja, no ya con letra manuscrita,

sino con distintas tipografías, lettererías sería la palabra exacta y no fuentes, de los distintos impresores de cuyos tórculos ha salido toda la producción editorial gallega. Justamente este punto, la dificultad del OCR para textos impresos en lenguas no inglesas, se ha destacado en la revisión elaborada por la Unión Europea dentro de su programa i2010, nacido como consecuencia no sólo de la necesidad de crear una industria de las Tecnologías de la Información y la Comunicación en Europa, sino del reto que supone el proyecto *Google Print*, al que recientemente se ha sumado, con otra iniciativa paralela, Microsoft en los Estados Unidos. Dentro de lo que es el programa Ingenio 2010 existe una línea estratégica denominada Avanz@⁸, uno de cuyos subprogramas es la elaboración de Contenidos Digitales, gestionado por la entidad privada de capital público Red.es. Creo que Galicia no debería quedar ausente de este proyecto y estoy seguro de que no será así.



<http://www.planavanza.es/>

8. <http://www.planavanza.es/>.

ORGANIZACIÓN DE RECURSOS DIGITALES

Los recursos digitales deben catalogarse, al igual que cualquier otra forma de producción intelectual recogida en un soporte, mediante su correspondiente normativa. Existen, por lo tanto, unas reglas de descripción específicas denominadas ISBD (ER), que ha venido a sustituir las antiguas ISBD (CF). También las FRBR⁹ jugarán un papel fundamental, así como el *Conceptual Reference Model*¹⁰, soporte conceptual básico para la realización de ontologías, definido por el Comité Internacional de Documentación del Consejo Internacional de Museos (CIDOC-ICOM), que cuenta con el apoyo del programa de Tecnologías para la Sociedad de la Información (IST) de la Unión Europea y que está en una fase muy avanzada de transformación en la norma ISO 21127. Existe un grupo de trabajo para la armonización de la normativa FRBR con CRM, que esperemos que de frutos próximamente.

En cualquier caso, una revisión global de la situación puede verse en *De re bibliografica galaeciana*¹¹ y existe una excelente colección denominada *Bibliofilia de Galicia*, que dirige Ignacio Cabano, y que recientemente se ha visto editada en soporte digital¹². Una recopilación sobre los recursos digitales sobre Galicia debería formar parte de los proyectos de visibilidad de las colecciones digitales. El Ministerio de Cultura recientemente ha puesto en marcha un proyecto cooperativo en el cual se relacionan los distintos proyectos de digitalización en España, incluyendo, lo que es muy inte-

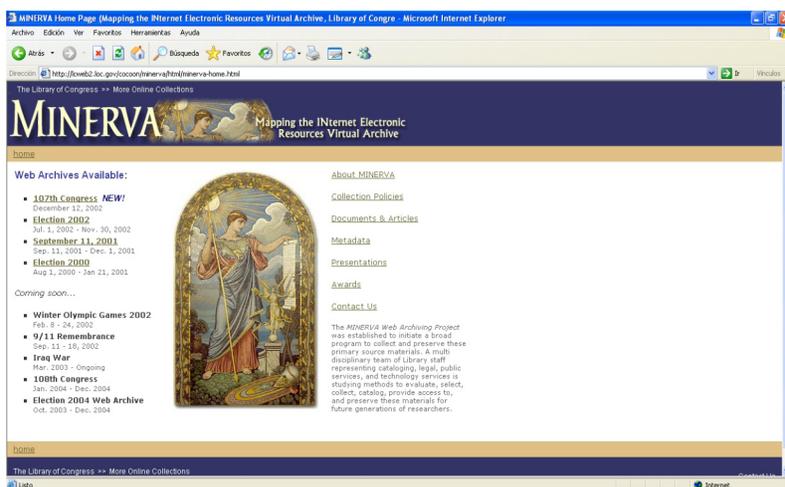
9. *Requisitos funcionales de los registros bibliográficos: informe final* / Grupo de Estudio de la IFLA sobre los Requisitos Funcionales de los Registros Bibliográficos; aprobado por el Comité Permanente de la Sección de Catalogación de la IFLA; traducción de Xavier Agenjo y María Luisa Martínez-Conde. – [Madrid]: Ministerio de Cultura, Secretaría General Técnica, [2004]. – 239 p. – ISBN 84-8181-213-7.

10. <http://cidoc.ics.forth.gr/>.

11. *De re bibliografica galaeciana, situación y perspectiva de la bibliografía en Galicia* / Ignacio Cabano Vázquez, Mercedes Fernández Valladares, Xavier Agenjo Bullón // *Cooperación, realidade e futuro: II Xornadas de Arquivos, Bibliotecas e Museos de Galicia: A Coruña, 24-26 de abril de 1997* / [organiza] ANABAD-Galicia. – [Santiago de Compostela]: Consellería de Cultura e Comunicación Social, 1997. – pp. 655-756.

12. *Bibliofilia de Galicia* (tomos 1 ao 23) 1ª ed. Santiago de Compostela, Xunta de Galicia, 2005. CD-ROM –(Patrimonio bibliográfico de Galicia. Serie facsimilar, n.º 2).

resante, los e-prints, así como un recolector OAI-PMH. Resulta claro, pues, que se debería dotar a los recursos digitales ya existentes en Galicia que se han ido mencionando tanto de repositorios que pudieran ser recolectados por proyectos tales como OAIster, Scirus o el del propio Ministerio de Cultura, así como al mismo tiempo se debería intentar una política de conservación de esos recursos digitales según la estrategia que ha dado origen a proyectos como PADI¹³, FEDORA¹⁴, MINERVA¹⁵ o Wayback Machine¹⁶. Este conjunto de acciones que vengo mencionando forman parte, a nivel bibliotecario, de los objetivos del denominado ICABS (*IFLA-CDNL Alliance for Bibliographic Standards*)¹⁷.



<http://lcweb2.loc.gov/cocoon/minerva/html/minerva-home.html>

13. Programa de la National Library of Australia denominado Preserving Access to Digital Information <http://www.nla.gov.au/padi/>.

14. <http://www.fedora.info/>.

15. Mapping the Internet Electronic Resources Virtual Archive, Library of Congress. <http://lcweb2.loc.gov/cocoon/minerva/html/minerva-home.html>.

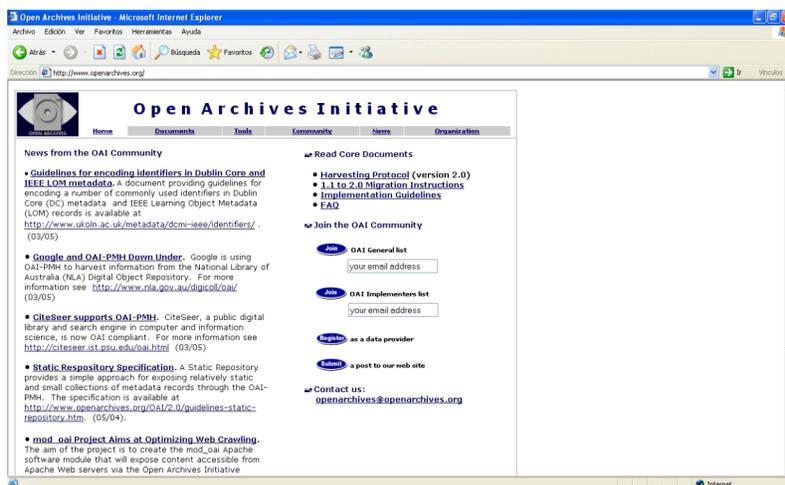
16. <http://www.archive.org/web/web.php>.

17. «Influencia del ICABS en el futuro digital de las bibliotecas» / Javier Agenjo Bullón, Francisca Hernández Carrascal. – En: *Boletín de la ANABAD*, ISSN 0210-4164, tomo 54, n.º 1-2, 2004, pp. 387-396. También accesible en <http://www.anabad.org/admin/archivo/docdow.php?id=121>.

PRINCIPALES PROBLEMAS Y RETOS

Está claro que, dadas las dimensiones de la web y su tendencia al crecimiento incremental, uno de los mayores problemas que se plantean es el de la visibilidad de los recursos digitales, así como la preservación y recuperación de este tipo de recursos, todo lo cual supone una mejora en la búsqueda y recuperación de esta información, que han de superar las limitaciones de los buscadores más importantes y que están en la mente de todos.

En principio, esas limitaciones están basadas en que fundamentalmente consultan páginas estáticas HTML, aunque algunos de ellos acceden desde hace poco tiempo a documentos PDF, y que también sólo buscan hasta un determinado nivel jerárquico y no son capaces de profundizar en la llamada «deep web», en la cual la información está organizada en bases de datos. Sin duda alguna, la solución más efectiva es la de recolección de metadatos mediante el protocolo OAI-PMH, tal y como se ha mencionado antes.



The screenshot shows a Microsoft Internet Explorer browser window displaying the Open Archives Initiative website. The browser's address bar shows the URL <http://www.openarchives.org/>. The website has a blue header with the text "Open Archives Initiative" and a navigation menu with links for Home, Documents, Tools, Community, News, and Organization. Below the header, there is a section titled "News from the OAI Community" with several bullet points. On the right side, there are sections for "Read Core Documents" and "Join the OAI Community" with input fields for email addresses.

Open Archives Initiative

Home Documents Tools Community News Organization

News from the OAI Community

- **Guidelines for encoding Identifiers in Dublin Core and IEEE LOM metadata.** - document providing guidelines for encoding a number of commonly used identifiers in Dublin Core (DC) metadata and IEEE Learning Object Metadata (LOM) records is available at <http://www.ukoln.ac.uk/metadata/dcmi-ieee/identifiers/> (03/05)
- **Google and OAI-PMH Down Under.** Google is using OAI-PMH to harvest information from the National Library of Australia (NLA) Digital Object Repository. For more information see <http://www.nla.gov.au/digicoll/oai/> (03/05)
- **CiteSeer supports OAI-PMH.** CiteSeer, a public digital library and search engine in computer and information science, is now OAI compliant. For more information see <http://citeseer.ist.psu.edu/oai.html> (03/05)
- **Static Repository Specification.** A Static Repository provides a simple approach for exposing relatively static and small collections of metadata records through the OAI-PMH. The specification is available at <http://www.openarchives.org/OAI2.0/guidelines-static-repository.htm> (05/04).
- **mod_oai Project Aims at Optimizing Web Crawling.** The aim of the project is to create the mod_oai Apache software module that will expose content accessible from Apache Web servers via the Open Archives Initiative

Read Core Documents

- **Harvesting Protocol (version 2.0)**
- **1.1 to 2.0 Migration Instructions**
- **Implementation Guidelines**
- **FAQ**

Join the OAI Community

Join OAI General list
[your email address]

Join OAI Implementers list
[your email address]

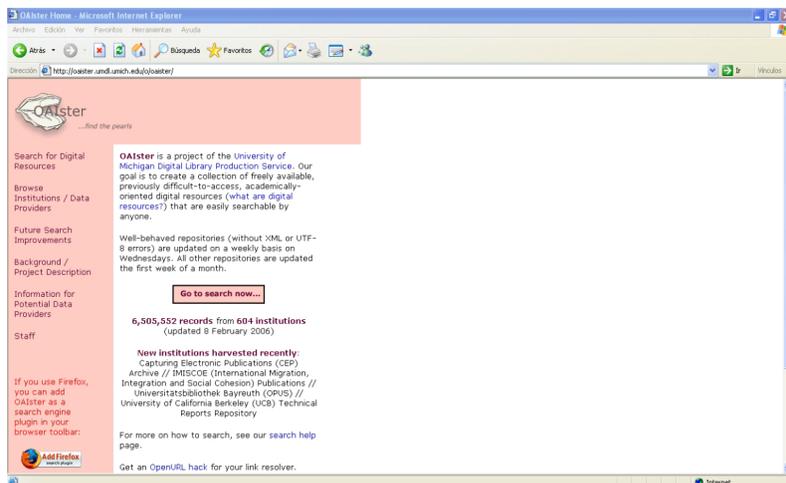
Register as a data provider

Submit a post to our web site

Contact us:
openarchives@openarchives.org

<http://www.openarchives.org>

El crecimiento de esta solución es verdaderamente considerable. Así, cuando preparé la presentación para la conferencia que pronuncié en el Consejo de Cultura Gallega, presenté datos –actualizados a 23 de noviembre de 2005– sobre OAster, que recogía 6 073 500 registros procedentes de 572 instituciones. Cuando actualizo estas páginas, el número de registros a fecha 8 de febrero de 2006 es de 6 505 552 registros de 604 instituciones, es decir, 432 000 recursos digitales nuevos en poco más de dos meses, lo que supone un aumento del 6,64%. Y no se olvide que no es ni con mucho el único recolector que existe. Scirus, dirigido más específicamente para recoger información digital en ciencia y tecnología, experimenta incrementos similares.



<http://oaister.umdl.umich.edu/oaister/>

Conviene destacar que el protocolo OAI-PMH supone un importante avance por su carácter de búsqueda sobre datos consolidados sobre la normativa Z39.50 ANSI/NISO o ISO 23950, no sólo por la complejidad de aplicación de este protocolo, sino por la falta de precisión en búsquedas simultáneas por distintos grados de aplicación de la norma. Este problema se está intentando solventar con la muy poco o nada utilizada en España ZING

(Z39.50 *International Next Generation*) y muy especialmente con el protocolo de búsqueda y recuperación SRW/U y, lo que es más interesante todavía, por la conjunción de todas estas normativas en un solo modelo. En relación con este aspecto de la recuperación de la información en la red convendría destacar el trabajo realizado por la Universidad Carlos III de Madrid, la Universidad de Salamanca y la empresa DIGIBIS, que, bajo la dirección de la investigadora principal, Francisca Hernández, ha llevado a cabo una completa revisión a nivel nacional e internacional del estado del protocolo OAI-PMH y que pronto estará disponible en la red.

El modelo OAI no resulta especialmente difícil de implementar. En realidad lo mínimo es la creación de repositorios estáticos¹⁸ y a continuación extraer de una base de datos mediante la conversión de registros de una estructura de metadatos que sea accesible a través de un servidor web en una dirección URL. Es decir, no es necesario catalogar en Dublin Core ISO 15836 directamente, sino que resulta muy sencillo mediante la correspondiente interfase de la que ya disponen algunos programas creados en España, generar los registros en DC. Es importante señalar que el DC sin cualificar es la estructura mínima necesaria, pero que, así mismo, es posible recolectar información en DTD y esquemas de XML para archivos como EAD¹⁹, el formato e-prints tan habitual en el mundo académico, y directamente en MARCXML Schema.

18. *Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting*. <http://www.openarchives.org/OAI/2.0/guidelines.htm>.

19. Los últimos dos catálogos publicados por el Archivo del Reino de Galicia exportan la información en EAD, por lo tanto, y una vez creado el correspondiente repositorio, serían así mismo recolectables.

Real Audiencia de Galicia, catálogo de preitos e expedientes de mosteiros [Recurso electrónico] / dirección e introducción, Gabriel Quiroga Barro; realización, Cristina Bañobre Fraga... [et al.]; diseño e tratamento da base de datos no Arquivo do Reino de Galicia, Rafael Lorenzo Durán, Luis Suárez Tajés. – [Santiago de Compostela]: Xunta de Galicia, [2004]. – 1 CD-ROM. – ISBN 84-453-3864-1.

El mismo proceso se ha llevado a cabo con el *Fondo documental del Pazo de Brandeso*, conservado en el Arquivo do Reino de Galicia y que se encuentra en fase de replicación.

En esta misma línea convendría mencionar otros proyectos como los que se citan a continuación y que cumplen con las funcionalidades descritas:

Acuerdo Universidad de Michigan + Yahoo

- Repositorio y Recolector OAIster accesible a través de Yahoo
- Libros, artículos, informes técnicos, imágenes, películas, discursos
- Difusión del contenido de proveedores (repositorios como LC, *New York Public Library*, Proyecto Gutenberg, etc.)

Proyecto mod_oai, financiado por la Andrew Mellon Foundation

- Desarrollo de software para transformar la información contenida en servidores Apache en repositorios OAI
- El módulo mod_oai se desarrollará bajo licencia pública (*GNU Public License*)

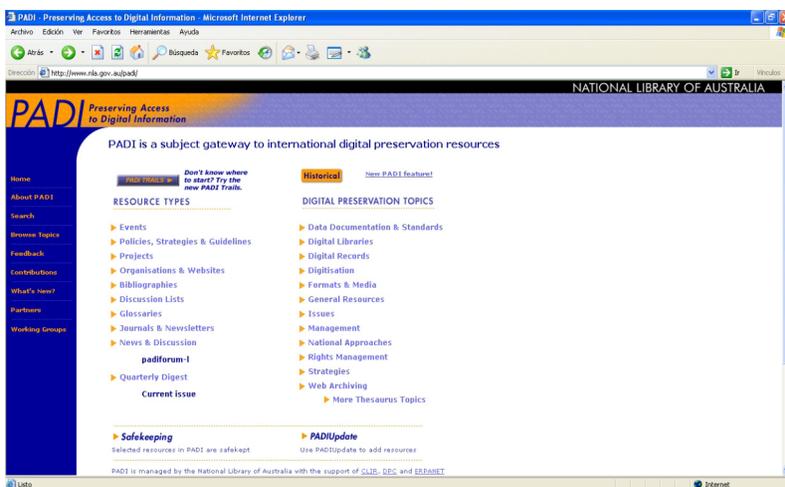
International Internet Preservation Consortium

- Bibliotecas Nacionales de Australia, Canadá, Dinamarca, Finlandia, Islandia, Italia, Noruega, Suecia, Inglaterra, Estados Unidos e Internet Archives.

MINERVA (Mapping the Internet Electronic Resources Virtual Archive)

- Promovido por la *Library of Congress*
- Conservar y difundir los recursos creados digitalmente e inexistentes en otro soporte físico

Acuerdo Google y Biblioteca Nacional de Australia (marzo 2005)



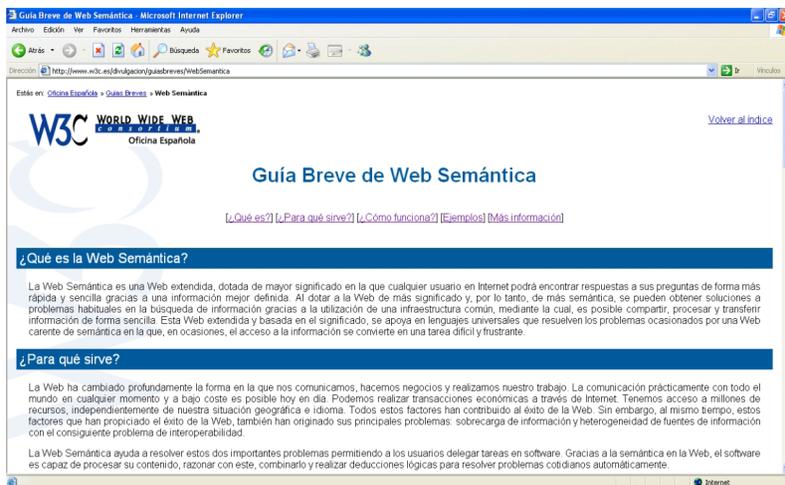
<http://www.nla.gov.au/padi/>

LA WEB SEMÁNTICA

En mi opinión²⁰, es necesario plantear y emprender un proyecto de singular importancia al que deberíamos denominar *Ontología del Patrimonio de Galicia*. Conviene recordar que las ontologías son el núcleo central del futuro estadio de la web actual a la que se denomina web semántica, en la cual las relaciones y enlaces de la información estarán basadas en conceptos y significados, es decir, en el conocimiento; y permitirá la superación de uno de los mayores problemas que la red presenta en la actualidad, la sobreabundancia de información que se ofrece al usuario en función de criterios no explícitos, lo que en la teoría de la comunicación se denomina ruido.

20. «De las bibliotecas virtuales a las ontologías y la web semántica» / Xavier Agenjo, Francisca Hernández. – En: *II Congreso Nacional de Bibliotecas Públicas: La biblioteca pública, compromiso de futuro*: Salamanca, 17-19 de diciembre de 2004. – [Madrid]: Subdirección General de Coordinación Bibliotecaria, 2004. – pp. 354-364. Además de la edición en papel y en disco, puede consultarse en Internet en la dirección del Ministerio de Cultura y dentro de la sección Travesía: http://travesia.mcu.es/documentos/congreso_2bp/3a_sesion/comunicacion01.pdf.

El objetivo es que, gracias a las relaciones conceptuales establecidas entre las distintas clases de información y entidades que constituyen una ontología, el usuario pueda navegar por la información de un modo mucho más eficiente, como si estuviera consultando una enciclopedia temática en la cual cada una de las entradas estuviera relacionada con el resto mediante unos vínculos especializados y perfectamente definidos, entre los que destacarían por ser más intuitivos, tanto las relaciones espaciales como las temporales. La suma de conceptos y relaciones consignados en una ontología reflejarán el conocimiento concreto que se dispone en una determinada área, en nuestro caso el Patrimonio Cultural de Galicia.



<http://www.w3c.es/divulgacion/guiasbreves/WebSemantica>

ONTOLOGÍA DEL PATRIMONIO DE GALICIA

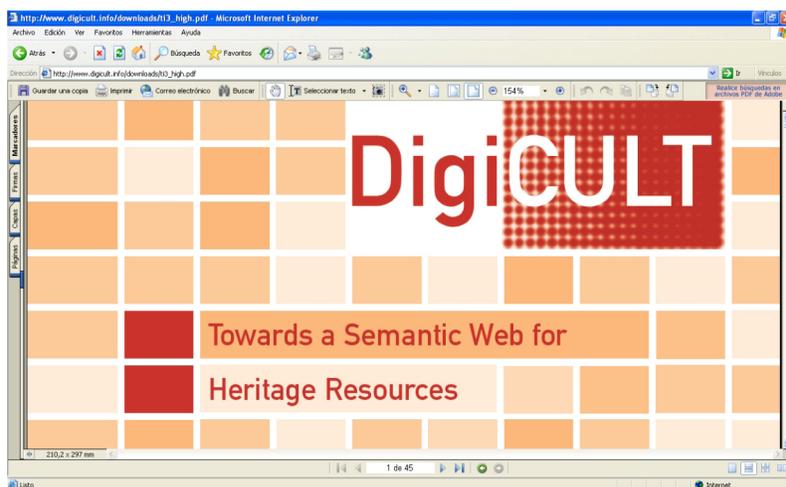
El proyecto sería importante por varias razones. En primer lugar, porque destaca la novedad radical de su planteamiento en cuanto a la aplicación de las nuevas tecnologías de la información al conocimiento y difusión entre todos los ciudadanos del patrimonio cultural. En segundo lugar, porque supone crear un instru-

mento de información extraordinariamente avanzado y potente. Y en tercer lugar, porque permitirá reunir de una forma estructurada toda la información existente sobre el patrimonio gallego, convirtiéndola no sólo en una fuente de información, sino también de conocimiento para todo tipo de usuarios. El proyecto tiene además tres objetivos prácticos concretos: la creación de una metodología de trabajo para la creación de este tipo de sistemas de información avanzados que suponen dentro de la Red de las ontologías; una recapitulación de los trabajos, estudios, análisis y compendios sobre el patrimonio gallego, y de las actuaciones realizadas sobre él, entre las cuales destaca indiscutiblemente las que ha llevado a cabo la Xunta de Galicia y otras instituciones como las universidades gallegas y el denominado tercer sector, es decir, las fundaciones; y la constitución de una fuente de información y conocimiento, que se prevé de gran impacto a medio plazo, que permita planificar y proyectar de una forma coherente el amplio conjunto de actuaciones vinculadas al desarrollo sostenible en el campo del patrimonio y el territorio.

El esfuerzo inicial se centraría en el análisis y establecimiento de los principios y planteamientos básicos que deben regir el desarrollo de un proyecto semejante. Se buscaría metodológicamente una adecuación total a la normativa internacional emergente en este campo, lo que concluiría con la adopción como modelo del *Conceptual Reference Model*, ya citado más arriba, así como del modelo conceptual definido por la *Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA)* en su influyente trabajo *Requisitos Funcionales para los Registros Bibliográficos (FRBR)*. Además, entre otras fuentes, es conveniente tomar en consideración los resultados y experiencias de proyectos europeos basados en la tecnología de las ontologías, como *VICODI: Visual Contextualisation of Digital Content* o *Finnish Museums on the Semantic Web*²¹, y las recomendaciones del programa europeo DigiCULT,

21. <http://www.vicodi.org/about.htm>.

que actúa como grupo de trabajo de referencia en los temas de patrimonio cultural digital, en su estudio monográfico *Towards a Semantic Web for Heritage Resources*²².



www.digicult.info/downloads/ti3_high.pdf

SELECCIÓN Y TRATAMIENTOS DE FUENTES DE INFORMACIÓN

De forma simultánea se deberían emprender también una serie de acciones dirigidas a establecer los conjuntos de información básicos que deberían incluirse en la ontología, que, en último término, no es sino una base de datos o conjunto de información particularmente bien ordenado. Afortunadamente, ya se ha citado gran parte de ese material producido o generado como resultado de las actividades de la Xunta, entre las que yo destacaría el Catálogo Colectivo del Patrimonio Bibliográfico Gallego, la Biblioteca Dixital de Galicia y la edición digital de *Bibliofilia de Galicia*, así como el amplio conjunto de iniciativas de las instituciones citadas ante-

22. www.digicult.info/downloads/ti3_high.pdf.

riormente, entre las que destaca la Biblioteca Virtual Galega²³ de la Universidade da Coruña y los trabajos de recopilación bibliográfica del Museo de Pontevedra²⁴.



<http://bvg.udc.es/index.jsp>

Para aumentar la operatividad del proyecto y empezar a disponer de un esbozo de prototipo podría tomarse la decisión de automatizar toda la información que citan Ignacio Cabano y Anxos García Sumai en el artículo reseñado, así como el resto de información digital referenciada, y, como puesta en práctica de las líneas metodológicas aquí expuestas, la información recogida por el Museo de Pontevedra y luego por el Instituto de Estudios Gallegos Padre Sarmiento del CSIC, utilizando un formato aprobado internacionalmente por la International Standard Office y del que también disponen otras fuentes potenciales de información para la Ontología como la Centro Superior Bibliográfico de Galicia, el ya citado Museo de Pontevedra, el Catálogo Colectivo del Patrimonio Bibliográfico de Galicia, así como el resto de información bibliográfica sobre estas materias recuperada en los catálogos de

23. <http://bvg.udc.es/index.jsp>.

24. «D. Xosé Filgueira Valverde e a Bibliografía de Galicia» / Ignacio Cabano Vázquez, Anxos García Sumai. – Museo de Pontevedra, 1997, 51:221-247. – ISSN 0210-7791.

las Bibliotecas Públicas del Estado, de la Biblioteca Nacional, de la Red de Bibliotecas Universitarias o, cuando sea posible, de la red de bibliotecas municipales que se gestionan actualmente con el programa Meiga. Toda esta información podría integrarse en un nuevo sistema de gestión documental con dos objetivos: uno, permitir la consulta en línea de información que antes sólo podía hacerse parcialmente en forma textual; y dos, lo que todavía es más importante, poder exportar toda esa información bibliográfica, utilizando formatos XML, datos y metadatos que son con los que se alimentaría informáticamente la estructura interna de la Ontología.

En la misma línea de trabajo debe contarse con un programa de gestión de la estructura de la Ontología. Entre los disponibles de código abierto destaca el programa especializado en la edición y mantenimiento de ontologías diseñado por Universidad de Stanford, denominado Protégé, que es ampliamente utilizado en todo el mundo para la creación de ontologías de cualquier tipo, y, en torno al cual, se ha creado una activa y abierta comunidad de especialistas en el desarrollo de las ontologías. También puede tomarse en consideración el software KAON (lo que es importante si se desea acudir a una financiación de la Unión Europea) desarrollado por la Universidad de Karlsruhe. Utilizando las funcionalidades de cualquiera de estos dos programas se puede estructurar un primer esqueleto de información, según el análisis conceptual al que se ha hecho referencia más arriba, e incorporar además alguno de los registros bibliográficos procedentes del proceso anterior o de distintas fuentes, como por ejemplo el Instituto Nacional de Estadística, para disponer de un completo *Nomenclator* que pudiera fijar la información geográfica.

De forma paralela, la actividad de selección de contenidos informativos para la Ontología se debe completar por la labor que debe realizar el equipo de mantenimiento del proyecto, cuya primera misión será realizar una primera selección de las fuentes de

información sobre el patrimonio de Galicia existentes en la red, de forma que fuera posible descargarlas en un momento dado en la estructura de la *Ontología del Patrimonio de Galicia*.

Además, se da la circunstancia de que dentro del proyecto Polígrafos²⁵ se está llevando a cabo la realización de *Antonio López Ferreiro Dixital*, que prácticamente toca todos los aspectos relacionados con el conocimiento producido por la cultura gallega hasta el siglo xx, como un primer esbozo de lo que podría acabar consultándose cuando se disponga primero del prototipo y posteriormente de la *Ontología del Patrimonio de Galicia*.

ESTUDIO DE VIABILIDAD

Con todos estos elementos (análisis conceptual, definición del esqueleto de la Ontología en el programa Protégé o KAON, conversión informática de los registros bibliográficos del Museo de Pontevedra, volcado de los mismos en un nuevo sistema de gestión bibliográfica que permita generar registros en lenguaje RDF, adopción de un modelo de recogida de datos y actualización de la información y selección de fuentes digitales sobre el patrimonio gallego) se debería afianzar la experiencia obtenida con la realización de un estudio de viabilidad que permitiera evaluar con precisión aspectos tales como el esfuerzo en personas/hora para el desarrollo del proyecto, la selección del software para el desarrollo de tratamiento de la información, la elección del sistema de gestión de la base de datos, así como las interfases entre los distintos programas (Protégé, KAON) y aplicaciones destinadas a una muy amplia gama de usuarios, que acabarán conformando la web de consulta de la *Ontología del Patrimonio de Galicia*.

Este estudio de viabilidad permitirá planificar, precisar costes y plazos de realización y comenzar a levantar el edificio sobre los planos, por así decirlo, que se han delineado en las fases previas. Se dispondrá, de este modo, de un orden de magnitudes, sin duda ya muy próximo a la realidad, sobre el coste final del proyecto, su duración, fases e hitos del mismo y una propuesta

25. <http://www.larramendi.es/bvirtuallarramendi.htm>.

de distribución de contenidos, para lo que se habrá tomado en consideración la legislación existente, puesto que las actuaciones sobre el patrimonio tendrán que atenerse a ella, se dispondrá de una primera maqueta que contendrá la mayoría de los elementos de la ontología, aunque referidos sólo a uno o dos partes del patrimonio gallego.

- Como resumen, y hablando en términos generales, las instituciones de memoria deben proseguir una línea de actuación centrada en proporcionar diferentes tipos de información a diferentes tipos de usuarios (investigadores, docentes, público en general); en crear servicios y contenidos para educación; en ofrecer un acceso completo a las colecciones (bibliotecas digitales, exposiciones virtuales, etc.); en mostrar cada uno de los objetos y sus descripciones (metadatos). Pero también en mostrar las relaciones entre los objetos (por medio de la interrelación semántica de los metadatos), para lo cual los archivos, bibliotecas y museos están en una posición inmejorable para lograr el potencial de la Web Semántica, trasladando los catálogos, tesauros, clasificaciones y taxonomías a esta web. La experiencia y acumulación de práctica les sitúan en una posición de confianza, con un conocimiento exhaustivo y fiable del dominio, pero dependerá del desarrollo de procesos de marcado sintáctico y semántico del contenido, de herramientas de análisis del conocimiento, de la adopción de lenguajes de representación del conocimiento (XML, RDF, OWL), del establecimiento de metodologías para definir y extraer el conocimiento de un dominio (construcción de ontologías).

CONCLUSIONES

A lo largo de esta comunicación se ha tratado de poner de manifiesto cómo se han llevado a cabo determinadas iniciativas en Galicia, conducentes al control bibliográfico, tanto corriente como retrospectivo y se ha efectuado un importante programa tanto de digitalización de todo tipo de materiales, así como de edición digital de repertorios fundamentales para el conocimiento del patrimonio bibliográfico, archivístico o incluso del museológico.

Se ha planteado que una de las mejores soluciones para lograr una mayor visibilidad de los recursos en la red, así como para recolectar informaciones digitales generadas por terceros, sería la adopción del protocolo OAI-PMH, lo que se reflejaría, por un lado, a través de la creación de una serie de repositorios estáticos y dinámicos y, por otro lado, mediante la implantación de un recolector que permitiese el rastreo en toda la red de informaciones referentes al patrimonio cultural gallego. Para la primera de las acciones bastaría con crear uno o varios repositorios sujetos a las directrices de la iniciativa OAI-PMH y dotar las catalogaciones existentes –de cómo mínimo– de una estructura de metadatos, que sería el Dublin Core sin cualificar. También se ha mencionado que idéntica actuación podría llevarse a cabo con los archivos, puesto que la recolección de metadatos contempla, así mismo, la EAD, en la que ya empiezan a editarse (y se mencionan algunos ejemplos) algunas de los fondos documentales del *Arquivo do Reino de Galicia*.

Por último, y ya con una mayor ambición, se propone la creación de una Ontología del Patrimonio Gallego (bibliográfico, archivístico y museológico, comenzando por el primero), en la que se esbozaría brevemente una metodología de actuación para llevar a cabo el proyecto. Conviene recordar que a lo largo del artículo se ha mencionado tanto algunos de los procesos ya realizados de digitalización del patrimonio bibliográfico gallego como la normativa necesaria para que esta digitalización alcance la máxima visibilidad en la red e incluso se estructure en formas complejas y avanzadas de estructuras de bases de datos, como podría ser una ontología, en la que se señalaría la existencia de programas estatales y comunitarios, es decir, Ingenio 2010 y, a nivel europeo, i2010, que no sólo recomiendan y requieren la normativa que se ha ido citando, sino que pueden proporcionar, al menos parcialmente, fondos para la digitalización y puesta en valor del patrimonio bibliográfico, de tal manera que éste sea accesible para todo el mundo desde cualquier sitio y en cualquier momento con el mayor grado posible de exhaustividad.