

DEL ALMACÉN A LA WEB: LAS COLECCIONES DIGITALES DEL ICC

MONTANER GARCIA, M^a CARMÉ - CAP DE LA UNITAT DE LA CARTOTECA DE CATALUNYA.
ROSET ARISSÓ, RAFAEL - RESPONSABLE DE CARTOTECA DIGITAL.

Resumen:

En octubre de 2007 la Cartoteca del Institut Cartogràfic de Catalunya puso online 5.000 mapas y fotos antiguas, libres para consulta y descarga pero con restricciones de uso, correspondientes a la digitalización de sus fondos antiguos. Era la culminación de un plan de digitalización empezado años antes y el inicio de una nueva cartoteca digital que va incrementando continuamente los fondos disponibles en la red.

Palabras clave: cartografía, cartography, cartoteca, map library, colecciones digitales, digital collections, mapas antiguos, antique maps.

1. INTRODUCCIÓN

El Institut Cartogràfic de Catalunya (ICC) cuenta, desde 1985, con una Cartoteca pública dedicada a recopilar cartografía antigua y moderna no sólo de Cataluña y del resto de España, sino también de todo el mundo. Uno de los objetivos, desde su fundación, ha sido poner a disposición del público mapas antiguos, tanto originales como facsímiles, y mapas modernos con una especial atención a la cartografía topográfica oficial a distintas escalas que van del 1:25 000 al 1:500 000 y más. En estos casi veintitrés años de funcionamiento, más de 300 000 hojas cartográficas que cubren todo el mundo, avalan este objetivo y ofrecen un servicio público en nuestra sede de Montjuïc en Barcelona. Esta documentación cartográfica se ha ido completando con una biblioteca técnica, un fondo de fotografías convencionales y diversos fondos documentales relacionados con temas territoriales. Toda la información puede localizarse a través de nuestros catálogos e inventarios consultables en internet: www.icc.cat/cartoteca. Un punto y a parte lo constituye la fototeca con más de 340.000 negativos de fotografías aéreas verticales, una parte significativa de los cuales está digitalizada.

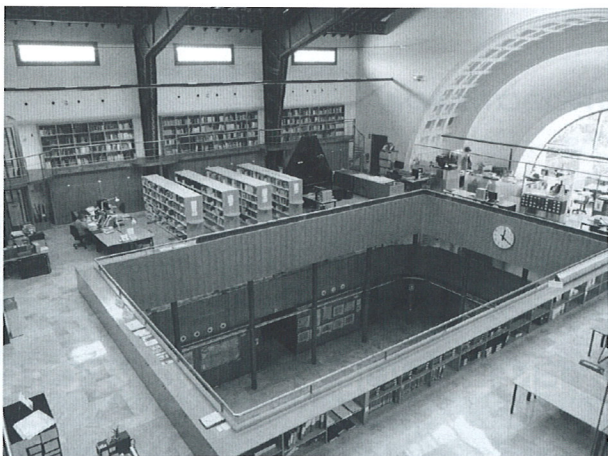


Figura 1: Instalaciones de la Cartoteca en la sede del ICC en Barcelona. En primer término está la sala de consulta pública. Al fondo, el escáner Metis al lado de la biblioteca técnica. El ICC ocupa un edificio construido con motivo de la Exposición Internacional de 1929 en la falda de Montjuïc.

Las nuevas tecnologías también han entrado con fuerza en el mundo de las cartotecas. Los catálogos automatizados de bibliotecas permitieron, sobretudo a partir de la década de 1980, la descripción de la documentación cartográfica dentro de los estándares bibliotecarios (Ibermarc, Catmarc) con adaptaciones específicas para el material cartográfico (ISBD (CM)). A inicios del siglo XXI, los avances en técnicas de digitalización, están revolucionando el mundo de las cartotecas. La visualización de cartografía digitalizada a través de Internet, está transformando las posibilidades de búsqueda, visualización y obtención de reproducciones de los usuarios. La visualización del mapa hace obvias las descripciones de determinados contenidos, tan habituales en los catálogos clásicos. De hecho, el mapa digital también se ha convertido en un campo de descripción dentro de los catálogos georeferenciados como base de situación y localización geográfica de cualquier elemento descrito. La irrupción y éxito de Google maps es el mejor ejemplo.

En el ICC, aunque las técnicas de digitalización entraron muy pronto a formar parte de los procesos de producción de cartografía, éstas no empezaron a aplicarse a su Cartoteca hasta inicios del año 2003. Este año, desde la Cartoteca empezaron a definirse los parámetros de digitalización, los flujos de producción, las nomenclaturas, los almacenes digitales, etc. que habrían de permitir dar los primeros pasos en el proyecto de digitalización de sus fondos. Unas primeras pruebas con un escáner DinA3 y unos primeros mapas escaneados con un escáner de rollo, nos permitieron iniciar un lento proceso que año en año iría en aumento. Éste mismo año, se pusieron en Internet las primeras imágenes de mapas digitalizados y al año siguiente, se añadieron URL a las fichas de catalogación de mapas del catálogo colectivo impulsado por el CBUC (Consorti de Biblioteques Universitàries de Catalunya) - en el cual participa la Cartoteca del ICC- y que permitió visualizar, desde esta plataforma, algunos mapas antiguos de Catalunya.

El 2007 ha sido el año de la consolidación del proyecto de digitalización de la Cartoteca en dos sentidos: en la captura de datos y en la visualización a través de la web. A principios de año, se dotó la Cartoteca con un escáner Metis DRS2A0, que permite escanear documentos de 60x90 cm. a 600 ppi hasta documentos de 120x180 cm. a 300 ppi. Este escáner de luz fría incorpora una mesa de presión que permite digitalizar sin arrugas mapas planos sin perjudicar el original, y su adquisición fue posible gracias a una subvención concedida en el año 2006 por la Secretaria de Telecomunicacions i Societat de la Informació de la Generalitat de Catalunya. De esta manera desaparecían los

problemas de formato y manipulación para una parte muy significativa de los fondos antiguos de la Cartoteca, en relación con el proceso de digitalización.



Figura 2: Equipo de digitalización Metis DRS2A0 del ICC, con capacidad para digitalizar originales de hasta 120x180 cm a 300 ppi.

En cuanto a la visualización, la adquisición de la aplicación ContentDM, producto de la empresa americana Dimema, ahora integrada en OCLC (una organización sin fines de lucro dedicada a facilitar el acceso a la información del mundo y reducir el ritmo de aumento de los gastos de la biblioteca) nos permitió poner a disposición de los usuarios a través de Internet, a finales del mismo año 2007, más de cinco mil imágenes con funciones de zoom y con campos de búsqueda alfanuméricos. Desde febrero de 2008, las imágenes pueden ser descargadas a alta resolución sin coste alguno. Podemos afirmar que la Cartoteca digital del ICC se ha puesto definitivamente en marcha.

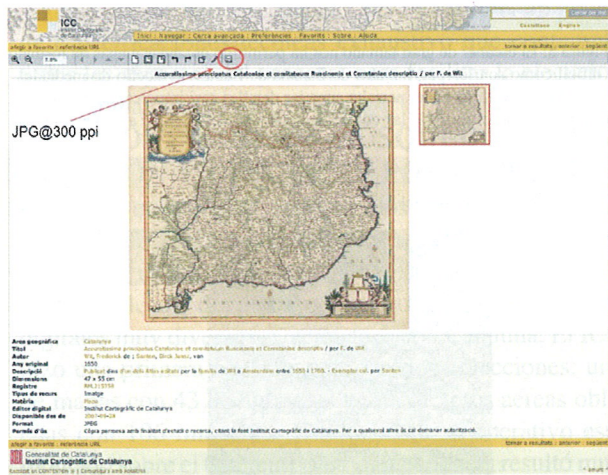


Figura 3: Todos los documentos cartográficos de la Cartoteca digital del ICC pueden descargarse gratuitamente a alta resolución. Sólo hay que pedir permiso para su utilización pública i/o comercial.

2. LOS CONTENIDOS DE LA CARTOTECA DIGITAL

Desde la web del ICC (www.icc.cat) se accede directamente a la página de la Cartoteca que está encabezada por la Cartoteca digital. Dada la gran cantidad de documentos y en espera de una futura georeferenciación, se han organizado colecciones temáticas, estructuradas básicamente en áreas geográficas:

Cada una de estas colecciones permite acceder a una lista de documentos, que pueden ordenarse con distintos cri-

terios y se previsualizan con una imagen miniatura. Una vez escogido el documento se accede a su visualización con las funciones de zoom. La imagen se acompaña con una descripción de la pieza basada en el estándar Dublin Core, en este caso con un perfil propio de la Cartoteca del ICC para cada colección.

Colección	Periodo	Número de imágenes 27/may/2008	Formato original
Mapas de Cataluña	s. XVII-XXI	684	Mapas
Mapas de España	s. XV-XX	1.116	Mapas
Mapas de Europa	s. XVI-XXI	1.102	Mapas
Mapas de África	s. XVI-XXI	123	Mapas
Mapas de América	s. XVII-XXI	216	Mapas
Mapas de Asia, Oceanía e islas del Pacífico.	s. XVI-XX	204	Mapas
Mapas del mundo y grandes regiones.	s. XVI-XX	111	Mapas
Vistas de todo el mundo	s. XVI-XX	974	Grabados
Fotoplanos de ciudades catalanas	1945-1949	389	Foto mosaicos
Miñutas municipales de ciudades catalanas	1914-1936**	865	Mapas
Archivo Gaspar de fotografías oblicuas de ciudades catalanas	1929	106	Fotografías aéreas oblicuas
Revista Catalana de Geografia.	1978-1996	3.607	Páginas de texto
Ediciones de guerra del MTN 1:50 000	1936-1939	275	Mapas

Tabla 1: Descripción de las colecciones digitales del ICC en fecha 27 de mayo de 2008.

Se puede navegar por las colecciones o se pueden realizar búsquedas precisas. Existe una búsqueda general por área geográfica - situada en el encabezamiento- que busca los resultados en todas las colecciones, y una búsqueda avanzada que permite hacer todo tipo de combinaciones.

La interfaz de consulta se ofrece en tres idiomas (catalán, castellano e inglés), y la búsqueda por área geográfica se modificó para tener doble entrada: el campo Dublin Core "coverage-spatial" se duplicó para almacenar las formas catalana y castellana del mismo topónimo con el propósito de conseguir el mayor número de resultados posibles sea cual fuere el idioma de la búsqueda. A su vez se añadieron funciones de sugerencia de cadenas alternativas de búsqueda, en los casos en que no se obtienen resultados, basados en los algoritmos de distancia de Levenshtein.

Se pueden descargar, sin coste alguno, los archivos a baja y alta resolución. En este último caso es necesario registrarse como usuario, igual que se hace con la descarga de los otros productos del ICC. El fichero digital del mapa se acompaña un archivo de texto con la catalogación del documento. Las copias solo pueden utilizarse para fines de estudio. Para cualquier otro uso es imprescindible solicitar autorización al ICC. De esta manera, un usuario remotamente conectado a la Cartoteca del ICC, no sólo puede buscar y visualizar los mapas sino que en el mismo proceso puede descargarse la imagen digital. El documento que se servía, y se continúa sirviendo, en soporte papel en la sala de consulta de nuestras instalaciones, ahora se sirve también en soporte digital al usuario virtual. El objetivo sigue siendo el mismo, los procesos se han diversificado.



Figura 4: Pàgina principal del gestor de colecciones digitales del ICC en Internet (www.icc.cat)

3. EL PROCESO DE DIGITALIZACIÓN

La digitalización se organiza a partir de la selección de conjuntos de documentos o “lotes” que tienen unas determinadas características comunes y que permiten definir unos mismos parámetros de digitalización. De esta manera se pretende optimizar al máximo el proceso. En dicha selección se tienen en cuenta criterios como la antigüedad y el valor patrimonial de la colección, el valor documental intrínseco y el interés que despierta en los usuarios, el formato y la homogeneidad del fondo, el estado de conservación y el estado de catalogación. También hay que tener en cuenta los derechos de propiedad ya que solo podrán publicarse en línea aquellos materiales libres de copyright o cuyos derechos sean de la propia institución. En la Cartoteca del ICC los lotes se han definido por formatos y por antigüedad, priorizando los documentos de reserva o aquellos con un estado de conservación considerado precario.

Todo el proceso de digitalización requiere necesariamente un manipulado cuidadoso de los objetos a cargo de un operador especializado, que disponga de las herramientas mínimas para su tarea: guantes de algodón blancos, grandes superficies de trabajo planas y libres a su disposición, un lugar seguro para almacenar los originales mientras dura el proceso de digitalización del lote correspondiente. A su vez el escáner, como equipo fotográfico que es, necesita de unos cuidados diarios, como la limpieza de las zonas de trabajo para evitar la aparición de cuerpos extraños (motas de polvo, ralladuras en el cristal, etc.) en las digitalizaciones. También necesita de unos cuidados periódicos como herramienta de precisión, referidos al conjunto óptico, la calibración de color, la longevidad de las lámparas de luz fría, las tolerancias mecánicas del equipo y la revisión general de la operativa.

Una vez establecidos los lotes de trabajo es necesario definir también las condiciones de digitalización, pues pueden variar de un lote a otro según los originales que lo compongan. En el caso del proyecto del ICC se estableció que el formato digital de preservación sería archivos TIFF neutros a 600 ppi, y que el formato de consulta y distribución sería archivos JPG factor de calidad 10 a 300 ppi. Debido a las limitaciones del escáner los originales de formato superior a 90x60 cm son capturados a resoluciones inferiores a 600 ppi, y por tanto actualmente los archivos digitales de preservación siguen siendo en formato TIFF

neutro pero con resoluciones que varían entre los 300 ppi y los 600 ppi.

En el caso de los archivos fotográficos la resolución óptica de captura que empleamos es de 2.032 ppi, pues tanto el escáner como el software lo permiten. Aunque parezca muy elevada esta resolución de captura es la que permite obtener luego tamaños de impresión de hasta DINA4 para originales fotográficos de 35 mm. Actualmente, y para los fondos fotográficos, estamos ensayando técnicas de superresolución para disminuir los tiempos de captura sin merma de calidad.



Figura 5: El escáner de gran tamaño nos ha permitido escanear mapas de formato difícil como son los provinciales de España de Francisco Coello de mediados del siglo XIX. El zoom permite apreciar detalles casi imperceptibles en el original. Como el resto de las colecciones pueden descargarse a alta resolución.

Figura 6: Un detalle del mapa de la provincia de Barcelona de F.Coello en su version digital

Tan importante como el equipo de digitalización es la estación de trabajo en donde se procesan los archivos digitales. Esta estación de trabajo debe ser una plataforma muy rápida, con una cantidad enorme de RAM (a partir de 2 GB), con mucho espacio libre en disco (mínimo 500 GB) y una muy buena conexión de red para el traslado de los archi-

vos originales de la captura a dispositivos de copia permanente que pueden estar alojados en un centro remoto. Hay que señalar que un original típico de 90x60 cm. digitalizado a 600 ppi produce un archivo digital de 800 MB.

Y que no es infrecuente digitalizar un original en dos o más partes que luego deben ser unidas en un archivo digital único, cuyo tamaño puede acercarse rápidamente a los 2 GB. Actualmente la Cartoteca digital del ICC ocupa aproximadamente 16 TB.

Junto al proceso intrínseco de digitalización está la catalogación de éstos archivos a metacatalogación. Si bien para cada uno de los lotes partimos de una catalogación descriptiva del documento (tanto del soporte como del contenido), una vez obtenido el archivo hay que catalogar este nuevo soporte. En el proyecto del ICC, es el propio operador de digitalización el que entra estos datos en una base ACCESS que después se implementará en el catálogo. Los campos básicos de metadatos para el proyecto de nuestra Cartoteca son los siguientes:

- Resolución
- Nombre de la Cartoteca
- Fecha de digitalización
- Restricciones
- Nombre del fichero
- Modelo del escáner
- Soporte digitalizado
- Observaciones

Un aspecto previsto en este proyecto es la externalización de algunos lotes de digitalización. En algunos casos (soporte fotográfico, etc.) se han hecho digitalizaciones en empresas externas a la propia institución. Para ello, se hace un estudio de los proveedores para evaluar aquellos que dispongan de los equipos necesarios pero también de personal especializado para llevar a cabo con éxito la operación. Para la correcta externalización de los trabajos es imprescindible haber redactado, para cada lote, un pliego de especificaciones técnicas exhaustivo que contemple todos los aspectos del proceso, para facilitar la ejecución de los trabajos por ambas partes.

4. EL SOFTWARE DE BÚSQUEDA, VISUALIZACIÓN Y DESCARGA POR INTERNET

El CBUC (cuya misión es mejorar los servicios bibliotecarios a través de la cooperación) invitó en el año 2006 al ICC a participar en el proyecto "Memòria Digital de Catalunya". Se trata de un depósito cooperativo de documentos digitales muy diversos relacionados con Cataluña. El ICC hizo una primera contribución con dos colecciones: una de mapas con 43 documentos y una de fotos aéreas oblicuas con 106 imágenes. Este depósito cooperativo está montado sobre el ContentDM. La experiencia resultó muy positiva por ambas partes.

Cuando la Cartoteca del ICC empezó la búsqueda de un software para poner sus fondos en Internet, la experiencia con la Memòria Digital de Catalunya, tuvo una influencia importante.

Después de evaluar distintas opciones, tanto de desarrollos a medida hechos con medios propios dentro del ICC, como de plataformas de código abierto, al final se optó por adquirir este producto que ya nos era conocido y con el que habíamos realizado una primera experiencia. Este gestor de colecciones digitales se caracteriza por estar desarrollado en PHP y tener una base muy grande de implantaciones en los ámbitos universitarios y bibliotecarios de

Norteamérica. La existencia de un número tan elevado de usuarios favorece el intercambio de conocimientos entre la comunidad, lo cual redundará tanto en la resolución de problemas, como en las mejoras que puedan añadirse al producto base.

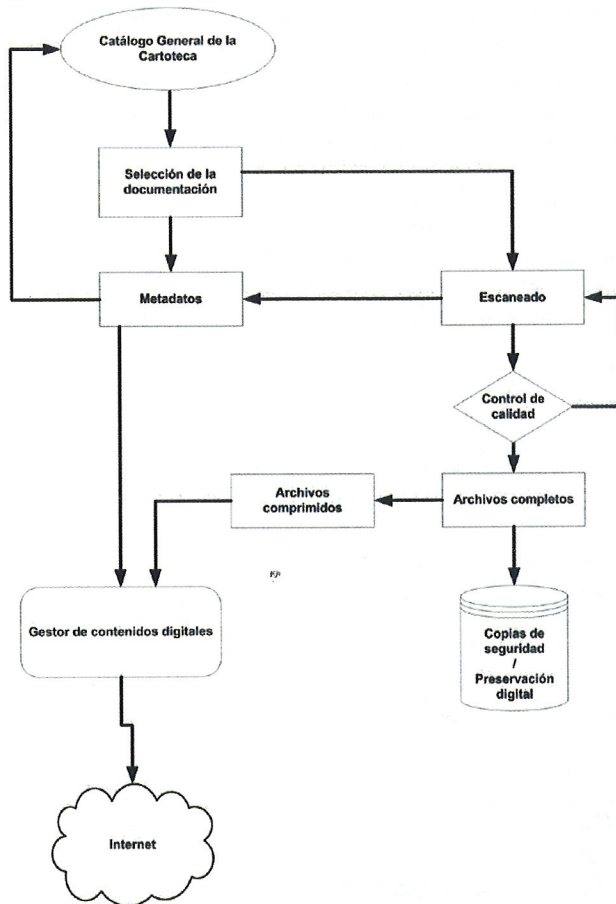


Figura 7: Flujo del proceso de digitalización de documentos que se sigue en la Cartoteca del ICC

La instalación en el ICC se hizo sobre servidores RedHat, con arquitectura Intel, dotados con capacidad de disco suficiente para albergar todos los documentos existentes. A partir de la instalación, y con la ayuda del distribuidor español del producto, se modificaron las plantillas de presentación para adecuar la visualización de las colecciones a la imagen corporativa del ICC. En este punto se hicieron las primeras pruebas de carga de documentos, y se desarrolló un proceso interno que permite la transformación de los metadatos provenientes de las bases de datos propias en metadatos expresados según el estándar Dublin Core (una norma de metadatos en línea interoperables para una amplia gama de propósitos y modelos de negocio) para ser cargados en el gestor digital.

Una de las ventajas de este gestor de contenidos es que puede ser alojado externamente en servidores fuera de la propia empresa, ya que para la gestión y creación de las colecciones emplea un programa de usuario instalado localmente, muy fácil de utilizar. Y la administración de las colecciones en el sitio web se realiza únicamente con el navegador de Internet.

5. UNAS PRIMERAS CONCLUSIONES

La importancia de la digitalización y de Internet en el mundo de la documentación es innegable.

Pero en el ámbito específico de la cartografía este hecho es

aún más evidente. Nuevos hábitos entre los usuarios empujan a las cartotecas públicas hacia nuevas plataformas para servir la información. Un primer paso lo constituye, sin duda, la digitalización de los documentos cartográficos. La Cartoteca del ICC inició en el año 2003 este reto que se ha visto consolidado con la puesta en marcha de su Cartoteca digital a finales del 2007. El proyecto hasta el momento ha resultado altamente positivo tanto por le número de visitas y descargas efectuadas, como por los comentarios que nos han hecho llegar los usuarios.

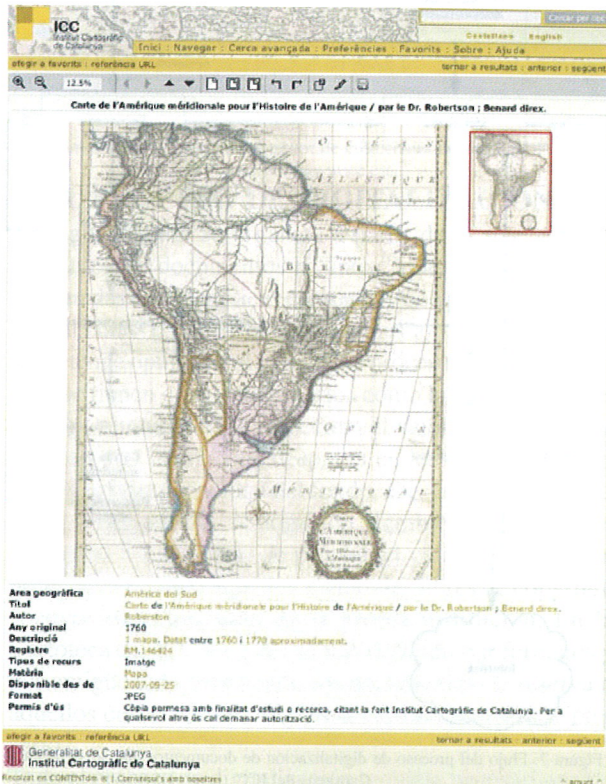


Figura 8: Las colecciones digitales de la Cartoteca no solo incluyen mapas de España sino también de todo el mundo. Entre ellos una buena colección de mapas antiguos del continente americano.

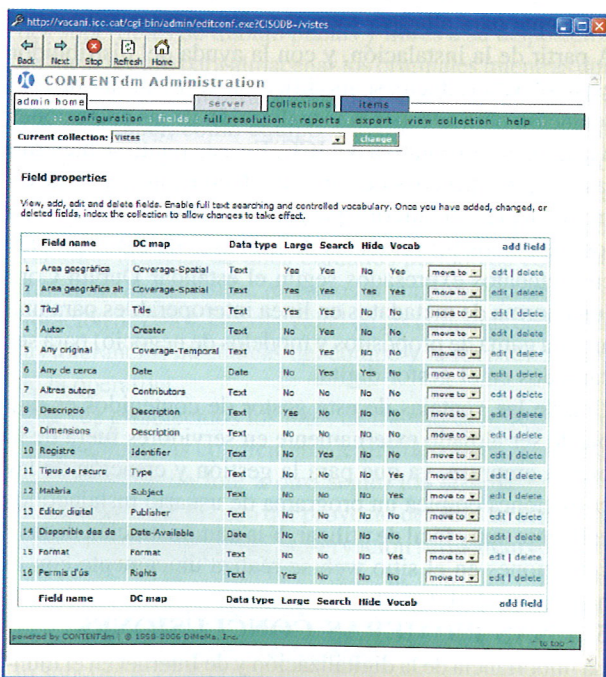


Figura 9: Ficha de administración de los campos Dublin Core de una colección en ContentDM

Puesto que la digitalización se acomete para facilitar la consulta de los fondos, disponer de un programa sencillo de recuperación de imágenes por múltiples campos es indispensable, y ofrecer la descarga gratuita de las imágenes de consulta ahorra muchas peticiones a los archiveros y bibliotecarios que pueden así verse liberados del trabajo de atención al público.

En el caso concreto de las cartotecas las herramientas de búsqueda y obtención de la información han cambiado, y están cambiando mucho más: debemos empezar a contemplar la georeferenciación como uno de los elementos clave de los catálogos de las cartotecas digitales. Si en este primer paso convertimos los campos de descripción MARC del documento original en campos de descripción Dublin Core para el documento en digital, debemos ir más allá y posibilitar la introducción en las cartotecas de las IDE (Infraestructuras de Datos Espaciales) de manera que podamos interoperar con todo tipo de catálogos con contenido cartográfico. El intercambio y la exportación de metadatos con las distintas IDE ayudarían a la descubierta de los contenidos por parte de los usuarios. Por otro lado, la recolección de datos y metadatos siguiendo el estándar OAI (Open Archives Initiative) ya nos permite consultar los contenidos desde otros portales como se ha demostrado en la colaboración con el proyecto europeo DigMap.

En el ICC hemos ya hemos empezado a trabajar en estos tres aspectos, que juntamente con otras iniciativas (social tagging, neogeography, crowdsourcing, etc) permitirán introducir la Cartoteca en la web de la próxima generación.

Referencias

Online Computer Library Center (OCLC)

<http://www.oclc.org>

Cartoteca del ICC (CTC)

<http://www.icc.cat/cartoteca>

Memòria digital de Catalunya (MDC)

<http://mdc.cbuc.cat>

Dublin Core Metadata Initiative (DC)

<http://dublincore.org/>

Infraestructura de Dades Espaciales de Catalunya (IDEC)

<http://www.geoportal-idec.net>

Open Archives Initiative (OAI)

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

Discovering our Past World with Digitised Maps (DIGMAP)

<http://www.digmap.eu>

Estàndards de digitalització. Elements mínims. Proposta del CBUC (actualitzat a juliol 2005)

<http://docs.cbuc.cat/fitxers/5digital/Public0507Minims.pdf>

Directrius per a projectes de digitalització de col·leccions i fons de domini públic, en especial els de biblioteques i arxius (juny 2006)

Col·legi oficial de Bibliotecaris-Documentalistes de Catalunya.

<http://www.cobdc.org/publica/directrius/index.html>

A Framework of Guidance for Building Good Digital Collections (3rd edition December 2007)

<http://www.niso.org/framework/>

Distancia de Levenshtein

http://en.wikipedia.org/wiki/Levenshtein_distance

Superresolució

<http://www.csic.es/prensa/noticias2006/02febrero/presentation.pdf>

<http://en.wikipedia.org/wiki/Superresolution>