

EVENTOS

VI WORKSHOP REBIUN PROYECTOS

DIGITALES. LAS BIBLIOTECAS DIGITALES

EN EL ESPACIO GLOBAL COMPARTIDO

2006



UNA EXPERIENCIA DE INTEGRACIÓN DE
REGISTROS DE CONTENIDO EN LOS CATÁLOGOS
AUTOMATIZADOS: EL PROYECTO ENRICHMENT



crue

Universidades
Españolas

Red de Bibliotecas
REBIUN

Una experiencia de integración de registros de contenido en los catálogos automatizados: el *Proyecto Enrichment*

Ricardo González Castrillo
Universidad Rey Juan Carlos

En diciembre del año 2005 se constituyó un Equipo de Trabajo formado por siete universidades -que son las que figuran en esta diapositiva [DIAP. 1] - con el propósito de llevar a cabo un proyecto, conocido con el nombre de *Proyecto Enrichment*. Culminaba, de este modo, una larga etapa de gestación cuyos orígenes se remontan al mes de octubre del año 2003, con los primeros contactos mantenidos entre SIRSI-IBERIA -llamada hoy SIRSI-DYNIX, empresa que, como es sabido, comercializa en España el programa de gestión UNICORN-, determinados librereros especializados y la Universidad Rey Juan Carlos. Tiempo después, el proyecto era presentado en la asamblea anual del Grupo Español de Usuarios de Unicorn celebrada en Ávila en junio del 2004, y en los meses siguientes de nuevo volvió a hablarse del tema en diversos encuentros con los directivos de esa firma comercial. Por último, tras estos pasos previos brevemente comentados y que aparecen en esta otra diapositiva [DIAP. 2], se acordó desarrollar el proyecto en la reunión de Navidad del referido Grupo el

pasado año, celebrada esta vez en Madrid, en la Universidad Carlos III, lo que supuso de hecho su comienzo formal. En la reunión, la Univ. Rey Juan Carlos hizo una nueva presentación del mismo informando, entre otras cuestiones, del procedimiento experimental adoptado por ella para la integración de registros de contenido. Y a la vista de su experiencia, se acordó que dicha Universidad actuase como coordinadora de un Equipo de Trabajo constituido al efecto, el cual ha venido funcionando con regularidad desde entonces [DIAP. 3].

La primera tarea que tuvo que abordar el Equipo fue la de definir, siquiera someramente, en qué consistía el *Proyecto Enrichment* y cuál era su propósito. Y en esta línea, se concluyó que básicamente su objetivo primordial había de consistir en la incorporación de ciertos contenidos a los registros bibliográficos del catálogo automatizado, con la finalidad de proporcionar al usuario, además de la información habitual de descripción y localización que facilita cualquier catálogo, otra adicional en la que se incluyeran la *imagen de la cubierta* de la obra, el *sumario* de la misma, y su *tabla o índice*. Así pues, son tres los contenidos que pueden aparecer añadidos a los registros catalográficos, de los cuales, casi siempre, están presentes la *imagen de la cubierta* y el *índice*. Y, con menor frecuencia, el *sumario*.

Una vez definido el propósito principal del proyecto, el siguiente paso del Equipo consistió en elaborar un *sistema de trabajo* común. Para ello, se concretaron, en primer lugar, las *condiciones de participación* en el proyecto, para fijar luego los *procesos* a seguir por sus miembros en la creación de registros de contenido, así como en la remisión de datos y en la corrección posterior de los mismos [DIAP. 4]. En el primer punto, las *condiciones* se establecieron muy genéricas con el fin de permitir la incorporación al proyecto

del mayor número posible de instituciones. Bien entendido que habían de cumplir, desde luego, un requisito imprescindible de carácter técnico: la utilización de UNICORN para la gestión bibliotecaria, con *i-link*, *i-bistro*, o alguna de sus variantes, como *interfaz* del catálogo automatizado. Cabe recordar, a este respecto, que no todos los miembros del Grupo Español de Usuarios UNICORN cuentan en la actualidad con esta herramienta, si bien muchos tienen ya previsto adquirirla a corto plazo. Se estableció asimismo la obligatoriedad de que los integrantes del Equipo aportasen un mínimo de 20 registros mensuales a la base de datos común. Luego veremos, al comentar las estadísticas, que muchos han rebasado con creces esta cifra.

Pero antes de continuar enumerando las restantes condiciones, quisiera hacer un inciso en la exposición para resaltar la expresión antes citada de *base de datos común*, porque precisamente es ésta una de las características más destacadas del proyecto, en cuanto está concebido como un verdadero *proyecto cooperativo*, en el que la información de contenidos generada por los miembros es compartida por todos ellos cuando existen registros bibliográficos coincidentes. Retomando de nuevo el apartado de las condiciones, pasamos a exponer las dos últimas establecidas por el Equipo con el carácter de recomendaciones a seguir. Y son, que los contenidos se hagan sobre obras editadas en castellano o en cualquiera de las lenguas oficiales del Estado español, y que la selección se realice, preferentemente, entre obras de reciente aparición en nuestro mercado editorial. Ambas condiciones encierran gran importancia y están plenamente justificadas dado que la empresa SIRSIDYNIX comercializa desde hace tiempo lo que podríamos llamar un *enrichment inglés*, es decir, los contenidos de libros vertidos al inglés, publicados principalmente en

Estados Unidos y en Gran Bretaña. Es lógico, pues, que teniendo cubierto ya los escritos en ese idioma, nuestro *enrichment español* aspire a hacer lo propio con los libros de esta lengua. El requerimiento referente a comenzar por las *novedades* parece ser también adecuado y conveniente. Sin embargo, al tratarse sólo de una recomendación como hemos apuntado, cada miembro ha optado por seguir su propio criterio. Y así, algunos han decidido trabajar sobre colecciones retrospectivas, o bien sobre materias o autores concretos.

[DIAP. 5] La etapa siguiente consistió en fijar un *procedimiento de trabajo* que fuese similar para los partícipes en el proyecto. El carácter cooperativo del mismo, ya resaltado, exigía disponer primeramente de una aplicación que pudiera advertir a los miembros, de forma inmediata, acerca de los registros de contenidos ya creados. A este respecto, quizá pudiera pensarse que la simple consulta de los catálogos bastaría para resolver esta cuestión. Y en verdad esto es cierto, pero tratándose sólo de registros cuya carga esté completada. Es decir, no es válido para los que se hallen en curso de preparación. De no hallar una solución, podría darse el caso de que varias universidades estuvieran haciendo al mismo tiempo un registro de contenido idéntico, con la consiguiente duplicidad de trabajo y los inconvenientes para su integración posterior. Por tal motivo, el Equipo hubo de adoptar una herramienta adicional que permitiera averiguar este dato antes de iniciar el trabajo. Y así creó una *base de datos de ISBN* que, mantenida por la Univ. Rey Juan Carlos y alimentada por las demás, informaba de esta cuestión. Su manejo es muy sencillo y proporciona una información de nivel básico pero suficiente. Por lo tanto, la primera operación a realizar consiste precisamente en acceder a dicha base de datos e introducir en ella el ISBN de la obra seleccionada. De no existir el código numérico, se añade éste a la base

datos junto con los relativos a la universidad que lo introduce, así como la fecha y hora en que se realiza esta incorporación.

Completada esta etapa, comienza el verdadero proceso de creación de registros de contenidos. La información de cada obra se agrupa en un fichero de texto plano convenido con la empresa SIRSI-DYNIX [DIAP. 6], un ejemplo del cual puede verse en esta imagen. Contiene cinco campos por libro - *ISBN*, *título*, *autor*, *sumario* e *índice* o *tabla de contenido*- que han de ser cumplimentados eliminando el que no conste. Ninguno de estos campos se considera obligatorio, excepción hecha del primero de ellos –el *ISBN*- por ser el que establece el vínculo del registro de contenido con el asiento del catálogo. Los campos de *título* y de *autor* se utilizan exclusivamente para identificar los contenidos de las obras y facilitar, al mismo tiempo, las tareas ulteriores de revisión pero no rectifican, en ningún caso, las etiquetas del formato catalográfico. Los de *sumario* e *índice* son, pues, en realidad los que se añaden a la entrada del catálogo como información adicional. Con referencia exclusiva a estos dos últimos campos, insistimos en que las etiquetas de identificación *summary* o *toc* deben ser suprimidas, como ya hemos dicho, cuando no contengan datos –caso frecuente en el primero, e improbable en el segundo- ya que, de no hacerlo así, aparecería el campo en blanco en el catálogo.

En el proceso de creación de contenidos, como apoyo al mismo, se emplean con asiduidad escáneres que contribuyen a agilizar esta labor. No obstante, y pese a su utilización generalizada, resulta casi imprescindible la revisión posterior de *sumarios* e *índices* con objeto de detectar posibles errores y subsanarlos antes de su procesamiento definitivo. Asimismo, estos equipos son también necesarios para la tarea de digitalización de la cubierta de la obra. Por

otro lado, es práctica habitual que tales trabajos los realice personal colaborador en las Bibliotecas, principalmente becarios-colaboradores formados en estas labores.

Una vez terminadas ambas fases, de elaboración y de revisión, los miembros del Equipo preparan, de manera individualizada, el material a remitir atendiendo a las siguientes normas: todos los contenidos de nueva creación han de agruparse en un único fichero de texto, en tanto que las cubiertas de los libros, también inéditas, se conforman en archivos independientes identificados por el ISBN.

Ahora bien, como sucedía que las universidades cometían erratas en los datos de estos contenidos, el Equipo hubo de establecer un proceso aparte para su corrección a fin de cuidar el aspecto de los registros en el catálogo. Tuvo que hacerlo así, además, ante las reticencias de SIRSI-DYNIX a suprimir de su servidor, sin más comprobación, versiones antiguas de contenidos ya existentes por otros más recientes. Por este motivo, se ideó un proceso especial para el tratamiento informático de registros modificados, diferente del seguido habitualmente con los nuevos. Y, en consecuencia, los miembros empezaron a diferenciar en sus envíos periódicos los registros de contenido en ficheros de texto separados, según fuesen actuales o no, e hicieron lo mismo con las cubiertas de los libros.

Tiempo es ya de explicar cómo se realiza la remisión de contenidos para su carga en el sistema. Mensualmente, antes del último día hábil de cada mes, todos los miembros mandan los archivos a una dirección corporativa de correo electrónico del proyecto, la cual es administrada por la Univ. Rey Juan Carlos. Y es esta misma universidad, en su calidad de coordinadora, la que organiza la

información en cuatro archivos –dos para ficheros de texto y otros dos para cubiertas- y los deposita en un servidor propio del que la compañía SIRSIDYNIX los retira automáticamente el quinto día de cada mes. La remisión final de un correo de confirmación de esta empresa acerca del proceso de carga de los contenidos –donde se detalla básicamente el número de *sumarios* e *índices* añadidos en su sistema y el número de *cubiertas* también incorporadas- cierra este procedimiento, pudiéndose ver desde entonces los resultados obtenidos en el catálogo automatizado.

Establecido, pues, el método de trabajo cuyos principales rasgos hemos venido comentando hasta ahora, el Equipo dedicó sus esfuerzos a la elaboración de una *página web* [DIAP. 7] con la intención de difundir el proyecto y buscar nuevas adhesiones al mismo, así como proporcionar una herramienta útil para los miembros. En la actualidad, se encuentra alojado este web en la página principal de la Biblioteca de la Univ. Rey Juan Carlos. Tras una breve definición del proyecto, con las universidades que inicialmente lo conforman y los resultados que se persiguen, se incluye luego una serie de documentos de interés, como las condiciones de participación, el procedimiento de trabajo y la presentación que del mismo se hizo en el Grupo Español de Usuarios de Unicorn en 2005, aspectos todos ellos ya comentados, además de la relación de los responsables de este proyecto en las distintas universidades y las estadísticas recogidas tras los procesos de carga de contenidos, desglosadas éstas por meses e instituciones.

[DIAP. 8] Con respecto a esta última cuestión, puede afirmarse, en verdad, que el incremento de la base de datos de contenidos ha sido mayor de lo esperado desde que en marzo de este año empezaron los envíos. Y hasta podría

calificarse de espectacular. Se ha pasado de 183 en dicho mes a 248 en abril, 281 en mayo, 404 en junio, 566 en julio, 274 en septiembre, y ¿??? en octubre. En números redondos, han sido ¿??? registros de contenido los creados en este tiempo. [DIAP. 9] En otras palabras: en estos ocho meses se ha conseguido sobrepasar holgadamente – en más de ¿???- las estimaciones de crecimiento que se habían hecho para todo el año los cuales, como se recordará, fueron de 1.680 a razón de 140 registros mensuales (20 por universidad).Y todo ello teniendo en cuenta, además, que no todos los miembros remitieron contenidos desde un principio.

[DIAP. 10] No obstante, el proyecto adolece de un grave inconveniente motivado por el hecho de que no tiene aún resuelto el espinoso asunto del *copyright*. Bien es cierto que hubo contactos previos con grupos editoriales españoles para solucionar esta cuestión durante los años de gestación, pero no lo es menos que tales esfuerzos no dieron el fruto esperado. La razón por ellos esgrimida fue el escaso interés que tendría la cesión de estos contenidos a las bibliotecas para su integración en los catálogos automatizados. Ni siquiera tomaron en consideración la oportunidad que podría suponerles al dar publicidad de sus producciones bibliográficas a través de este medio. Pese a este obstáculo, presente todavía, los miembros del Equipo decidieron en la asamblea del 2005 iniciar el desarrollo del proyecto. Sabido es que, en el mundo anglosajón, y, más concretamente, en alguno de los proyectos de enriquecimiento de contenidos que se están llevando a cabo hoy día en organismos como la Library of Congreso por el Equipo BEAT, como el D-TOC u ONIX-TOC, los contenidos son aportados por las propias editoriales o por empresas de servicios. Y es, precisamente, esta

segunda vía la que, en los momentos actuales, procura seguir el Equipo para solventar tal situación.

Ejemplo de ello es la reunión que el Equipo mantuvo en el mes de julio con SIRSI-DYNIX y el Taller Digital de la Universidad de Alicante, en respuesta al ofrecimiento de esta última para participar en el proyecto. De hecho, había ya realizado una presentación en la última reunión del Grupo Español de Usuarios de Unicorn en la que sentó las bases de esta futura colaboración. Su propuesta consistió, básicamente, en intentar obtener los permisos de las editoriales para la difusión de los contenidos, aprovechando los contratos generales de la empresa con el gremio de editores, y realizar además el proceso de digitalización de tales contenidos. Su solicitud al Equipo de un listado de las editoriales cuyas obras eran adquiridas por las universidades con mayor frecuencia, fue cumplimentada por medio de SIRSI-DYNIX al Taller Digital de la Universidad de Alicante, adjuntándole asimismo una carta del Equipo en la que se expresaba el deseo de colaboración.

Con independencia de cuál sea finalmente el resultado de estos contactos, lo que no puede cuestionarse es que, a la larga, la integración de registros de contenidos en los catálogos automatizados será una realidad debido, entre otras razones, a la mejora sustancial que supone para el usuario esta información sobre de los documentos de las colecciones de la biblioteca, inserta en los catálogos automatizados. Con una simple ojeada a los registros de contenido, podrá conocer si el libro o los libros en cuestión le serán de utilidad o no. Todavía más: la presencia de la cubierta escaneada, lejos de ser un embellecimiento vano, le permitirá identificar con mayor exactitud y rapidez tales obras en los fondos de libre acceso de las bibliotecas. Y aún más: la codificación de estos

contenidos en lenguaje HTML, a semejanza de lo que ocurre con los proyectos en curso de la Library of Congress, le posibilitará recuperar esta información en Internet, superando la limitación de que esta información no pueda ser buscada en los catálogos automatizados. Ojalá el futuro no nos defraude y frustre las esperanzas y las ilusiones que los integrantes del Equipo de Trabajo del *Proyecto Enrichment* han puesto en su realización.

Como colofón de esta exposición, finalizaremos la misma revisando un ejemplo de integración de contenidos en el catálogo.

Ejemplo: ISBN 8430931635

La familia y el cambio político en España

Julio Iglesias de Ussel

[Está en URJC y UC3M]