



- INICIO
- ZONA DE NOTAS
- NÚMEROS ANTERIORES
- SOBRE EPI
- DISTRIBUCIÓN/BASES DE DATOS
- SUSCRIPCIONES
- PUBLICIDAD
- INFORMACIÓN AUTORES
- AGENDA
- IWE TEL
- THINK EPI
- COPYRIGHT
- EQUIPO DE LA REVISTA

- CONTACTO

Junio 1993

Integración de catálogos en el Csic

Por **Teresa Malo de Molina**



La Red de Bibliotecas del CSIC inicia su automatización real en 1987 con el programa *Aleph* sobre un ordenador *Cyber* de *Control Data Corporation*. En 1989 se instala un VAX 6300 de *Digital* en Madrid, dedicado exclusivamente a la automatización de las bibliotecas y a las bases de datos del Csic, y en febrero de ese mismo año, se pone en marcha un VAX 8500 en Sevilla para constituir el nodo andaluz. Durante 1991 se instalan tres *MicroVAX 3300* en Barcelona, Valencia y Vigo,

constituyéndose así la actual arquitectura informática de la red, con 5 nodos destinados a mantener los respectivos catálogos locales.

Desde el principio se pensó como una red en estrella en la que cada nodo alojara el catálogo de los fondos de las bibliotecas de su zona, pero contando siempre con un catálogo colectivo global soportado en el nodo de Madrid, que fue el primero en la informatización y el que cuenta con el mayor número de registros. Esto implica, obviamente, la integración periódica de los registros soportados en los catálogos locales a este nodo central.

El primer nodo activo después de Madrid fue el de Sevilla, el cual, en julio de 1990 envió a Madrid la primera cinta con 3.000 registros de monografías (el catálogo colectivo de revistas ya estaba terminado por esas fechas). El personal del entonces *Programa de Informatización de Bibliotecas del Csic (Pribic)*, hoy *Unidad de Coordinación de Bibliotecas*, imprimió un listado y fue confrontando uno a uno todos los registros con el catálogo de Madrid para distinguir los coincidentes y los nuevos. Terminada la selección, estos últimos se cargaron en batch en el catálogo colectivo de Madrid.

El sistema era, como cabe suponer, lento y costoso, y evidentemente no resultaba rentable previendo el gran crecimiento que tendrían los catálogos locales, pues en ese momento ya se proyectaba la creación del nodo catalán y del valenciano. Además, a finales de 1990, gracias al proyecto "Biblioteca sin fronteras" auspiciado por la *Sociedad Estatal para el Quinto Centenario*, se informatizó íntegramente la biblioteca de la *Escuela de Estudios Hispanoamericanos de Sevilla*, aumentando el catálogo andaluz en cerca de 50.000 nuevos registros, que también había que integrar en el catálogo colectivo de Madrid.

Integrar automáticamente

El enorme número de registros obligó a buscar otras vías más rápidas y que supusieran menor carga de trabajo para la Unidad de Coordinación, pero que al mismo tiempo garantizaran la coherencia del catálogo colectivo impidiendo la duplicación de registros.

Así, a principios de 1991 se encargó a la Univ. de Jerusalén, la creadora de *Aleph*, un programa que permitiera filtrar las cargas masivas de registros en batch, comparando los registros de los conjuntos a integrar.

Se pensó que los dos campos más unívocos serían sin duda el *isbn* y el título, pero interesaba que la decisión final siguiera siendo personal, ya que, como es sabido, las normas del *isbn* no siempre se aplican exactamente (p. ej.

variantes de edición que no se traducen en cambios del *isbn*) y además muchos registros carecen de este campo. Interesaba que el programa *aparcara* los posibles duplicados en un fichero de trabajo manipulable.

Recibido el programa en mayo de 1991, se hizo una pequeña prueba con 293 registros de Sevilla, de los cuales el programa *aparcó* 92 como posibles duplicados. Éstos se listaron y se comprobaron manualmente, obteniéndose 63 verdaderos duplicados y 29 nuevos registros.

En ese momento el programa se limitaba a separar los registros cuyo *isbn* o título existieran ya en el catálogo colectivo, pero en el caso del título no ignoraba la grafía, de modo que sólo consideraba título encontrado aquel que estuviera exactamente igual escrito en el fichero existente de títulos, con las mayúsculas, minúsculas y acentos correspondientes. Como las variantes de grafía son abundantes en este campo, la realidad es que permitía la integración de registros duplicados por el simple hecho de un acento mal puesto o una mayúscula ausente.

Además, el fichero de registros *aparcados* tenía un formato diferente que impedía la carga en batch de aquellos que, tras la oportuna comprobación, resultaran nuevos registros a integrar, lo que obligaba a catalogar en el catálogo colectivo copiando los registros del listado impreso. Este problema se solventó a partir de febrero de 1992, cuando gracias al programa *Decnet* (*Digital Equipment Corp. Network*) de interconexión de ordenadores *Digital*, se tuvo la posibilidad de conexión transparente entre los catálogos de Madrid y Sevilla (v. IWE-3, p. 9) y de trasvasar nuevos registros de cada carga, directamente del catálogo de Sevilla al de Madrid. El uso de *Decnet* permitió además que en lugar de enviar cintas, Sevilla pudiera transferir los ficheros electrónicamente. Para esas fechas ya se habían integrado también registros de Barcelona y Valencia y, en estos casos, se había dejado pendiente la carga de nuevos registros hasta que estuviera en activo la conexión transparente con cada uno de estos nodos, cosa que hoy es ya una realidad.

De todas formas seguía presente el problema del escaso filtrado de los títulos, por lo que se encargó a la Univ. de Jerusalén que perfeccionara el programa para que ignorara las diferencias de grafía y, a ser posible, las palabras vacías iniciales (artículos, preposiciones, etc.) Así mismo se le solicitó que el fichero de registros *aparcados* tuviera el formato necesario para la carga en batch de los registros (esto sólo exigía modificar la presentación de la información).

A finales de 1992, la Univ. de Jerusalén envió el programa corregido que cumple lo indicado arriba, aunque en el *isbn* no ignora los guiones, con lo que no se pueden evitar los errores de tecleo en este campo. Ahora el formato del fichero de registros rechazados o *aparcados* ya es manipulable, de tal modo que basta con comprobar sobre listado si se trata de verdaderos duplicados o no. Si es así, se hacen las oportunas actualizaciones en el catálogo colectivo y, una vez corregido el listado completo, se trabaja directamente en el fichero con un editor borrando los registros duplicados y cargando después en batch únicamente los nuevos.

De este modo se han integrado ya en el catálogo colectivo de Madrid 62.734 registros de Andalucía, 16.935 de Catalunya y 4.360 de Valencia. El porcentaje aproximado de registros rechazados es del 20% en Andalucía, el 30% en Catalunya y el 38% en Valencia, resultando un porcentaje aproximado de duplicados reales del 20% en Andalucía, el 19% en Catalunya y el 27% en Valencia. Lógicamente, el número de registros rechazados y de duplicados es cada vez mayor debido al aumento del catálogo colectivo, que en la fecha actual cuenta ya con 308.782 registros.

Teresa Malo de Molina. Jefa de Coordinación Profesional. Unidad de Coordinación de Bibliotecas del Consejo Superior de Investigaciones Científicas (Csic)

Jorge Manrique, 27. 28006 Madrid.

Tel.: +34-1-585 44 43; fax: 564 42 02

Enlace del artículo:

http://www.profesionaldelainformacion.com/contenidos/1993/junio/integracin_de_catlogos_en_el_csic.html