



GACETA

Academia Colombiana de Ciencias Exactas, Físicas y Naturales

ISSN 0123-0654

Volumen I, Número 8, págs. 1-3

Agosto de 1997

Las bibliotecas digitales y la conservación del patrimonio cultural

VÍCTOR ALBIS

Departamento de Matemáticas y Estadística, Universidad Nacional de Colombia

Queremos en este artículo analizar, aunque sea brevemente, el futuro de la conservación del patrimonio cultural de la humanidad, en general, y, en particular, el del caso colombiano. La primera parte, relativa a lo que actualmente se hace en el mundo con las nuevas tecnologías, se ha adaptado del trabajo de MICHAEL LESK citado en la bibliografía, cuya lectura completa recomendamos a los interesados. En la segunda parte, comentamos someramente el *status artis* de la conservación de este patrimonio en Colombia, y presentamos un proyecto parcial, actualmente en marcha, de recuperación y conservación digital del *patrimonio científico nacional*.

1. ¿Qué pasa en el mundo? En los últimos años el mundo conoció la culminación del gigantesco proyecto de la *Biblioteca Nacional de Francia*, cuyo nuevo edificio alberga más de 22 millones de libros en 395 kilómetros de estantes. Desde el punto de vista de la conservación, el patrimonio cultural de Francia estará garantizado, en condiciones microclimáticas óptimas por muchísimos años. Pero se hizo algo más: 110.000 volúmenes que cubren gran parte de la historia y la cultura francesas, se pueden leer ahora, sin tocarlos, desde centenares de computadores situados en el moderno edificio. Este hecho es una muestra de una tendencia de las grandes y más importantes colecciones de material bibliográfico existentes hoy en día: la **digitalización**. Por ejemplo, se espera que, para el año 2000, el 91% de la *Biblioteca Nacional de Francia* y 5 millones de libros de la *Biblioteca del Congreso de los EE. UU. AA.* estén digitalizados usando el medio más barato existente, la *escanografía de documentos*. Este procedimiento tiene algunas desventajas, siendo la principal que sólo permite *mirar* pero no *buscar* en un texto procesado en esta forma. Otra desventaja es la enorme cantidad de memoria que consumen. Los procedimientos de reconocimiento óptico de caracteres (es decir, aquellos que producen texto digitalizado a partir de imágenes), llamados también OCR (*optical character recognition*) ofrecen una fidelidad del 99% en texto llano (en el mejor de los casos, una docena de errores por página), pero en otro tipo de texto, como fórmulas matemáticas o químicas desplegadas, los errores pueden llegar al 70%. De hecho, los métodos de reconocimiento óptico de caracteres en uso tiende a confundir una fórmula, con índices, exponentes y fracciones, con varias líneas distintas, con resultados caóticos. En cualquier caso, se deben corregir manualmente estos errores, si queremos usar los textos digitalizados con confianza. Pero las revisiones manuales son lentas y onerosas. Por otra parte, todo archivo electrónico que pretenda reemplazar un texto impreso, debe incluir los dibujos y fotografías originales y, nuevamente, los actuales sistemas de OCR tienden a confundirse ante la presencia de ilustraciones. Afortunadamente han aparecido técnicas que resuelven este problema y usadas con éxito, por ejemplo, por la *American Chemical Society*, lo que ha permitido sacar 400.000 figuras, fórmulas y diagramas de 428.000 páginas de las revistas especializadas en ciencias químicas para incluirlos en las ediciones electrónicas de los *Chemical Abstracts*. Otro proyecto interesante es el *Journal Storage Project* (los interesados pueden visitar la siguiente

dirección electrónica: URL: <http://www.jstor.org/>, para ver una demostración), financiado por la *Andrew W. Mellon Foundation*, que intenta poner en línea las 10.000 revistas académicas más importantes del mundo, con la colaboración económica de las universidades norteamericanas.

En este proceso de digitalización de las bibliotecas, subsisten todavía otros problemas técnicos, entre los cuales mencionamos el de la estandarización de los formatos (protocolos), pero que estamos seguros tendrán soluciones apropiadas en un futuro cercano. De otra naturaleza, pero muy importante y delicado es el problema de los derechos de autor y propiedad intelectual. La compañía IBM pagó más de un millón de dólares en regalías y derechos de autor para producir su CD-ROM conmemorativo de los 500 años del descubrimiento de América.

Veamos ahora algunas de las ventajas que ofrece la digitalización de las bibliotecas. En primer lugar está la *protección de documentos frágiles y preciosos* que hasta al momento estaba garantizada microfilmándolos. La ventaja de su digitalización, además de su protección, es la posibilidad de hacerlos circular electrónicamente a través de la red de Internet o en CD-ROM, lo que los hace *accesibles simultáneamente en diferentes lugares del globo*. Otra ventaja, que sabrán apreciar en su justa medida los bibliotecarios, es la *simplificación de los aspectos operativos cotidianos en las bibliotecas tradicionales*. Finalmente, los actuales *problemas de almacenamiento* (en particular, la construcción de edificios adecuados par la conservación de los materiales bibliográficos tradicionales y los costos de inventario y conservación) disminuirán en el futuro.

Antes de terminar esta parte, creemos conveniente señalar lo siguiente: a) Como casi todo lo que se imprime actualmente tiene su origen en un archivo digital, *en pocos años el gran volumen de la producción intelectual de la humanidad, a partir de 1990, estará digitalizado o podrá digitalizarse*. b) Los usuarios se acostumbrarán cada vez más a usar sólo aquello que encuentran en línea. Esto producirá un *desafortunado desuso de todo aquello publicado antes de 1990* cuya digitalización, dados los costos y la legislación existente en casi todas partes sobre los derechos de autor, es una tarea de proporciones ciclópeas. c) *La longevidad de los archivos electrónicos no es una realidad, en cambio la obsolescencia tecnológica sí lo es*. Esto obligará a los bibliotecarios y archivistas a cambiar de formatos en pequeños lapsos de tiempo, aunque se espera que, sin perder demasiada información, pasar de un formato a otro no sea muy difícil.

2. ¿Qué estamos haciendo y qué podemos hacer en Colombia? En Colombia los proyectos de protección y conservación del patrimonio cultural e histórico del país se basan actualmente en la microfilmación de documentos. Entre ellos debemos mencionar el del *Archivo General de la Nación* y los de algunas universidades. No existe hoy, que conozcamos, un proyecto basado en la digitalización de documentos. Por otra parte, existen varias importantes colecciones como la de libros publicados en Colombia en los siglos XIX y anteriores que reposan algunos en los húmedos sótanos de la *Biblioteca Central de la Universidad Nacional de Colombia*, o su archivo general, que están en peligro de desaparecer ante la incuria y desgreño de las autoridades competentes.

Evidentemente, además de conseguir o construir locales adecuados para la conservación óptima de los originales, la solución más barata puede estar en la microfilmación y sistematización de la información contenida en los documentos. Sin embargo, el manejo de los microfilmes continuaría siendo el mismo que se le da actualmente a los impresos en papel, y estaríamos desaprovechando las ventajas que su digitalización ofrece para hacerlos circular electrónicamente a través de la red de Internet o en CD-ROM, y, por lo tanto, *accesibles, pero intocables, simultáneamente en diferentes lugares del globo*.

Circunstancias económicas y de diversa índole, conocidas por todos, nos previenen de proponer megaproyectos de digitalización del patrimonio cultural e histórico colombiano. Sin embargo, es posible empezar este proceso de manera gradual, a nivel público y privado, con proyectos específicos dimensionados a capacidades e intereses puntuales. Dentro de este orden de ideas, hemos propuesto a

la *Universidad Nacional de Colombia* y, por su intermedio, a COLCIENCIAS, continuar el *proyecto bibliográfico sobre la producción colombiana en ciencias exactas desde la Colonia hasta nuestros días*, incluyendo en esta nueva etapa la digitalización de los documentos pertinentes. El nuevo nombre del proyecto es el siguiente: *Sistema de información de la producción matemática colombiana*. El producto final será, pues, una **base bibliográfica**, de carácter analítico e interactivo, acompañada de los textos, digitalizados, de los siglos XIX y anteriores, puesta en una serie de CD-ROM y eventualmente en el servidor del Departamento de Matemáticas y Estadística de la Universidad Nacional de Colombia, en Santafé de Bogotá. Este proyecto se enmarca en el sub-programa *Bibliografía Matemática Colombiana* del *Programa sobre historia de las matemáticas en Colombia*, iniciado hace ya varios años.

Bibliografía

MICHAEL LESK, *Going digital*. Scientific American.