

THE REAL DEATH OF PRINT

Despite clashes with publishers over copyright, Google's plan to make millions of books available online is turning the tide for efforts to digitize the world's literature.

Andreas von Bubnoff tracks the demise of the printed page.

Vishwas Chavan travels a lot. An informatician based at the National Chemical Laboratory in Pune, India, he collects data on what types of animal live where in India to enter into a biodiversity database. Yet the specimens he hunts for have neither fur nor feathers, but yellowing pages and ageing dustjackets.

Much of the information Chavan seeks is in old, out-of-print tomes that are scattered around the world; about 2,500 of the 7,000 books he has unearthed were written in the first half of the nineteenth century. To find them, Chavan has spent years trailing around libraries. He dreams of the day when books such as these are scanned and made available as digital files on the Internet.

Chavan and other digitization visionaries paint a future in which books no longer gather dust on shelves, but exist as interconnected nodes in a vast web of stored literature, all accessible at the click of a mouse. So instead of hunting for specific books, scholars could search for specific information, customizing searches to suit their needs.

A few years ago, Chavan's dream seemed little more than a castle in the air. True, a number of mostly volunteer-driven or publicly funded projects had been scanning books and making them freely available on the Internet. But most efforts were limited.

In December 2004, the Internet search-engine company Google announced plans to change that. It said it would scan millions of books from five major libraries: the university libraries of Oxford, Harvard, Stanford and Michigan, and the New York Public Library. The announcement energized other organizations in the United States and in Europe, which soon unveiled similar plans to

scan and catalogue millions of books.

The move to digitize books is set to transform the worlds of publishers, librarians, authors, readers and researchers. Obscure specialist titles could find new readerships; librarians and information specialists will have to develop tools to catalogue and navigate this labyrinth of data; and authors and publishers may soon have to start thinking in digital dimensions, just as website designers and writers already do.

Bloody revolution

But revolutions are rarely bloodless and this one could soon get ugly. In the United States authors and publishers are squaring up against Google for a legal fight over copyright. Opinion is divided over whether the scanning projects being implemented by companies such as Google and Amazon (see graphic opposite) will hand control of the world's literature to private enterprise — and, if so, what this could mean. And with several independent scanning projects under way, it is still not clear how much of the information will be freely avail-

able, or where and how it can all be coordinated and accessed.

The idea to digitize books and make them available online has been around since the Internet's inception in the early 1970s. When the US Declaration of Independence was typed in and sent to everyone on a computer network on the night of 4 July 1971, it marked the birth of Project Gutenberg, the first book-digitization venture.

Since then, the project's 20,000 volunteers have scanned or typed in about 50,000 out-of-copyright books, says its founder Michael Hart, who works in the basement of his home in Urbana, Illinois, and, like the project's volunteers, for free.

Projects such as this are driven by the idealistic desire to make knowledge and literature freely accessible to all, but also by the benefits of having book collections easily searchable. "Being able to find it online is pretty much the same as having it online," says David Weinberger of the Berkman Center for Internet and Society at Harvard Law School in Cambridge, Massachusetts.



Assets such as searchability have prompted the National Science Foundation (NSF) in Arlington, Virginia, to get involved in an open-access enterprise called the Million Book Project. This is an international scanning effort with many participants, including Carnegie Mellon University in Pittsburgh, Pennsylvania.

Since the project began in 2002, about 600,000 out-of-copyright books have been scanned, although only about half of them are currently available online (see graphic). The scanning takes place in India and China, with books being shipped there temporarily from libraries around the world.

Made to fit

Searchability is also the main driving force behind commercial plans to scan books, including texts whose copyright has yet to expire. For example, if their products have been digitized, online booksellers can allow customers to search within books and browse a few pages before deciding to buy. In the United States, with the publisher's permission, Amazon puts searchable digital data from mostly copyrighted books online. Amazon says that several hundred thousand books are currently available for searching.

Amazon also offers the option of purchasing e-books and e-documents on its website, which can be viewed after downloading them to a portable reading device (see 'Will flexible screens be the end of paperbacks?'). The company expects these services to drive additional sales. Its 'search inside the book' feature increases sales by 8%, the company says. Scientific publishers, such as the US National Academies Press also see increased print sales when they allow their books to be viewed online.

But Google doesn't mention money in its announcement that it plans to make the contents of millions of copyrighted books searchable as part of its Google Book Search project. Its spokesman, Nate Tyler, says Google's motivation is to include literature that is currently only available offline in its mission to make information universally accessible. But the possibility that the company could gain financially from the move has raised hackles among US authors and publishing organizations.

"Google is in a class by itself because of the quantity of money and the level of centralization."

— Daniel Greenstein

In the autumn of this year, the Authors Guild and the Association of American Publishers filed a lawsuit against Google for copyright infringement. They complained that Google hadn't asked them for permission to scan copyrighted books.

Google has obtained the go-ahead from publishers to include some copyrighted works as part of its Book Search project, but not all. It argues that it does not need to seek permission for every book, because what it plans to do is permissible according to the 'fair use' exception of US copyright law. This allows copying for uses such as teaching, scholarship or research.

Google will, for example, not make the full text available, but only show 'snippets' of text around the search results if a book is still copyrighted. The company says that people are more likely to buy or borrow a book if they can search it this way, adding that the snippets are similar to the card catalogues found in

libraries. But Paul Aiken of the Authors Guild in New York City argues that the act of scanning the works is copyright infringement no matter how the texts are used.

The outcome of the lawsuit will depend on the courts' decisions over how the concept of fair use applies in the age of digital books and the Internet. Meanwhile, the rest of the scanning world is watching from the sidelines, and being careful to scan only books that are out of copyright, or to obtain the publisher's permission before scanning anything.

Google's plan has shaken up the digital-book world in other ways too. For one thing, many believe that its size and resources mean Google can pull off this feat — so large-scale repositories of digital books seem a more realistic and immediate prospect than ever before. Google has also galvanized its competitors, both public and private (see graphic) to redouble their efforts, and has placed a question mark over the future of libraries and librarians.

"I think Google is in a class by itself because of the quantity of money and the level of centralization," says Daniel Greenstein, librarian of the California Digital Library in Oakland, California. "Google has paved the way, created the appetite for this kind of activity, and anxiety on the part of libraries and publishers."

Out with the old

But Michael Gorman, president of the American Library Association, says he is not worried that libraries could become obsolete. As well as providing access to books, they serve as a place for people to meet and study, he says. And librarians' expertise in information management will still be needed. "We are not worried about our own jobs," agrees Dennis Dillon,



C. DARRIN/N. SPENCER

Will flexible screens be the end of paperbacks?

"It's nice. It's the size of a paperback book. It's very light — I actually forgot I had it in my bag."

The future of reading, according to Theresa Horner, director of e-book operations for the publisher Harper Collins, lies in neat electronic devices such as the Sony Librie she totes in her handbag. Although some bookworms might balk at the thought of reading a novel on a screen, it might not be long before portable electronic books revolutionize the book world in the same way that Apple's iPod changed the music scene.

Almost any device can be used to read an e-book. Nick Bogaty, executive director of the International Digital Publishing Forum (formerly the Open e-Book Forum) in New York thinks that the "vast majority" of people reading e-books now are doing so on their PDAs or smartphones. The small, handheld devices that people use to organize their lives can act as useful stores for reference books, or hold novels that while away a long commute. Fans of this approach say they stop noticing the screen size



when gripped by a good plot.

There are also devices that attempt to recreate a more familiar reading experience. Garth Conboy, president of eBook Technologies, recounts how his company's devices, which are like small tablet-PCs with simple page-turning buttons, are supplied to about 15,000 students at a military college loaded with their course material. This saves the college

money in printing, and has the advantages of online material.

But the device that aficionados drool over is the Sony Librie. Launched in Japan in April last year, it is similar in size, shape and weight to a book, and is the first device to have a paper-like display.

The display uses technology developed by E Ink in Cambridge, Massachusetts, and works by reflecting light, rather than emitting

light like an LCD or LED screen. The pixels are microcapsules containing particles of black and white ink with opposite charge. An electric field pushes one or other colour to the surface, producing a print-like effect. The device only gobbles energy when the display is refreshed. So the lifetime of batteries in such an e-paper system is measured in page-turns, rather than hours.

And the future? It's likely to be flexible. In September, Philips unveiled the first working prototype of an e-reader with a rollable display, aimed not just at e-books, but also at news reading and Internet browsing. In the 'Concept Radius' a grey-scale screen measuring about 13 centimetres wide, again using E Ink technology, curls up inside a slimline packet measuring 10 × 6 × 2 cm.

It might be worth waiting for. "We plan to start our production by the end of 2006, so I anticipate products coming into the market in early 2007," says Hans Driessen, a spokesman for Philips.

Jenny Hogan

associate director of the research services division of the University of Texas libraries at Austin. "The job is changing, which makes it even more fulfilling than it was before."

But Gorman is worried that over-reliance on digital texts could change the way people read — and not for the better. He calls it the "atomization of knowledge". Google searches retrieve snippets and Gorman worries that people who confine their reading to these short paragraphs could miss out on the deeper understanding that can be conveyed by longer, narrative prose. Dillon agrees that people use e-books in the same way that they use web pages: dipping in and out of the content.

Best of both worlds

Having a mix of both e-books and real books could be the answer. A mix would certainly help solve that perennial headache for libraries — the lack of shelf space and cost of keeping physical books. Ensuring that some libraries always keep a physical copy of a particular work means that they will be available through inter-library loans for readers needing a real book, adds Dillon.

Some of them are already dispensing with hard copies. The University of Texas at Austin, for example has about 10,000 copyrighted books and 300,000 out-of-copyright works that are available only as e-books, says Dillon.

Another person to be energized, but also alarmed, by Google's move is Brewster Kahle,

founder of the Internet Archive, a non-profit organization in San Francisco that archives web pages and other digital files. Although Google has never indicated that it plans to claim ownership over its digitized material or charge for search access, Kahle doesn't want to leave digital books entirely in the hands of private enterprise.

That's why, in October, he announced the formation of the Open Content Alliance (OCA). This aims to build a permanent

"Science is moving incredibly fast, and scanning old books is a complete waste of money."
— Matthias Ulmer

archive of multilingual digitized text and multimedia content, which, as far as possible, will be freely accessible.

Like the Million Book Project, the OCA will scan out-of-print books; the first few are already available online. The alliance hopes to rival Google's project in terms of scale; among the groups helping to finance its scanning efforts are Yahoo and Microsoft. Some libraries who were reluctant to join the Million Book Project for logistical reasons have signed up to the OCA. "We did consider the Million Book Project, but we were hesitant because we wanted to avoid shipping overseas," says Tom

Garnett, assistant director for Digital Libraries and Information Systems at the Smithsonian Institution Libraries in Washington DC, a contributor to the OCA.

For taxonomists such as Chavan, the OCA is perhaps the most interesting scanning project so far. Eight museums including the Natural History Museum in London have formed the Biodiversity Heritage Library Project, which will collaborate with the OCA to scan about one million volumes of biodiversity literature, much of it out of copyright.

But Matthias Ulmer, a German publisher who helped launch an e-book initiative by the German Publishers and Booksellers Association, thinks that scanning old books is "a complete waste of money". "Science is moving incredibly fast, even in the field of taxonomy," says Ulmer. Earlier this year, his association announced an initiative whereby some 100 German publishers are considering digitizing about 100,000 newly published books by 2006. Publishers will take their own digital raw data and place them on a network of their own servers. Scientists and others will then be able to access the books for a fee.

With 2005 seeing the birth of so many digitization projects, it might not be long before Chavan can realize his dream of hunting for new specimens from the comfort of his armchair.

Andreas von Bubnoff is an intern in Nature's Washington office.