

NewsBreaks
& Conference Reports

EBSCO
INFORMATION SERVICES



Years of Experience.

Experience, service, content, solutions.
Let's discuss your needs today.

Gale Group to Digitize Most 18th-Century English-Language Books, Doubles InfoTrac Holdings

by **Barbara Quint**

June 17, 2002 — Apparently they only count in hundreds at Gale Group: hundreds of sources, hundreds of clients, and now hundreds of years. Gale Group has announced a mammoth 20-million-page project that will bring to the Web most books published in the English language during the 18th century. Proclaimed as “the most ambitious single digitization project ever undertaken,” it reflects cooperation by The British Library and other leading research libraries. In another announcement, Gale has more than doubled its InfoTrac holdings with the integration of 5,400 titles from ingenta. This makes a total of more than 9,000 electronic periodicals.

“We own the 18th century,” boasted Mark Holland, publisher in Gale’s U.K. office (and it sounded like he meant more than just the title of Gale’s prospective digital edition of *The Eighteenth Century*). When finished, the project will include the full-image text of 150,000 English-language titles published between 1701 and 1800. Gale plans to complete the project in time to put the product on the Web beginning in June 2003.

Gale’s Primary Sources Microfilm operation already has over 12,000 reels of microfilm covering the 18th-century literature—this project will scan and digitize all the pages. An OCR (optical character recognition) program, tweaked to accommodate 18th-century printing conventions, will then generate a text-searchable database. When complete, searchers should have highlighting of search terms and downloadable MARC records, plus metadata tapping the full text of title and content pages and direct access to all illustrations.

“This library is fundamental to the creation and understanding of the modern Western world,” said Holland. “*The Eighteenth Century—Complete Digital Edition* makes research far more convenient and far faster. However, just as important is the rich functionality of the database. It will permit new research and teaching opportunities to a greatly expanded community of students, teachers, and historians worldwide.”

The Eighteenth Century—Complete Digital Edition will be published in subject categories and released over a 3-year period. The initial release of History and Geography will occur in mid-2003. It will represent around 20 percent of the total titles, according to Holland. After that, segments should follow in the order of Literature and Language; Social Science; Religion and Philosophy; Science, Technology, and Medicine; Law; Fine Arts; and Reference categories. Libraries can subscribe to modules or to the entire collection.

Of course, one could expect considerable problems in dealing with such ancient text. Gale will try to use the standards set by the Text Creation Partnership, based at the University of Michigan. According to Holland, maintaining consistent full-text searching features may entail some additional re-keying. In time, Gale hopes to add links to move back and forth among chapters.

It's too early for any fees to be set; however, Gale has established its price structure. Holland indicates that the company will offer two main options. Libraries can purchase a collection as published for permanent retention (no endless license fees) or they can subscribe annually. Holland believes the latter approach will have great appeal to libraries or academic institutions with short-term interests or limited project funds.

Gale has been aggressively expanding its collection of digital editions of prized archives. It already has a collection of early English newspapers. The agreement with The British Library comes on the heels of one with *The Times* of London to publish a complete digital edition of that newspaper's archives from 1785 through 1985. The World War II-era issues went online recently. The company is also working with the Winston Churchill archives to microfilm a million documents spanning the statesman's life.

Of course, one major issue occurs here, particularly with the 18th-century material. Isn't all this material public domain? What's to stop someone from buying one set and then pouring it out onto the Web for all to use, like a Project Gutenberg? Not much, according to Holland. He noted that a library could choose to purchase the set, rather than subscribe through a license, and then open the content stored on its servers to the open Web.

In fact, Holland speculated on some advantages that might emerge. Clearly the OCR text generation from digitized images will result in lots of errors, particularly for detailed, even nit-picking linguistic research. If users tapped into an 18th-century collection on the open Web, downloaded the material, then corrected the database, Holland wondered if it might lead users to send corrected copy back to Gale for correcting the master record. As a matter of fact, Holland said the idea for such a "virtuous circle" came up at some focus groups held by Gale.

ingenta and Gale's InfoTrac

A partnership between Gale and ingenta [see "Gale, ingenta Partner on InfoTrac Plus" at <http://www.infotoday.com/newsbreaks/nb020121-1.htm>] has come to fruition. In North America, Gale has launched InfoTrac Plus, a product that combines scholarly electronic journals from ingenta and Gale into a 9,000-title collection. ingenta contributes over 5,400 scholarly and academic electronic journals, combining them with Gale's 4,200 full-text titles.

"ingenta's research shows that libraries want one access point to online journal content, instead of the myriad of Web sites they're faced with currently," said Andrea Keyhani, ingenta's chief operating officer. "The combination of ingenta's journal content and Gale's InfoTrac system forms the gateway to the most complete online, full-text journal content available in the marketplace."

In an exciting development, InfoTrac Plus will offer a pay-per-view option that lets users buy an individual digitized article when the library does not subscribe to the title. When libraries do subscribe, the system will try to verify "appropriate copy" status. It will also allow users to set maximum costs. When users choose the pay-per-view option, however, the option will only apply to the 5,400 journals from ingenta, according to Keyhani. Access to Gale Group sources remains subscription-only.

The partnership debuts with Gale's InfoTrac OneFile and Expanded Academic ASAP, two premier

InfoTrac products. InfoTrac already has users in more than 100 countries and supports InfoMarks, a persistent URL that allows researchers to save searches. Beta tests have been conducted at major universities, including Harvard, Virginia Polytechnic Institute, SUNY–Geneseo College Libraries, University of Texas–Austin, and the University of Minnesota–Twin Cities.

Barbara Quint is contributing editor for NewsBreaks, editor in chief of Searcher, a columnist for Information Today, and a longtime online searcher. Her e-mail address is bquint@mindspring.com.



[Previous NewsBreaks](#)



[Information Today Inc. Home Page](#)