

DH20015608 E1

# A Project Seeks to Digitize Thousands of Early English Texts

BY GOLDIE BLUMENSTYK

**A**CADEMIC LIBRARIES are working with a major corporate partner, ProQuest Information and Learning, to create digital editions of 25,000 early English texts.

A new database will hold fully searchable versions of works by Bacon, Galileo, Newton, and others, produced from 1475 to 1700.

ProQuest (formerly Bell & Howell) owns and sells microfilmed copies of the works—some 125,000 texts in all. In 1998, the company produced an electronic version of the archive by digitizing images of its pages. But the product that ProQuest sells, Early English Books Online, is searchable only by title and keyword. The new database will be a fully transcribed digital edition that will be fully searchable.

About 150 libraries are being recruited to finance the project by paying \$10,000 a year for five years. So far, 50 have pledged.

ProQuest is also contributing to the project, promising a \$1 match for every \$5 from libraries. It hopes that the fully searchable digital edition will make its own digital archive a more attractive product for libraries.

## A MODEL FOR OTHER PROJECTS

Academic librarians, led by those at the University of Michigan at Ann Arbor and the University of Oxford, say their involvement will help ensure that the new digital files meet the technical and scholar-

ly standards that researchers expect. The unusual project—making business partners out of parties that are sometimes at odds with each other over pricing and access issues—could eventually be a model for future collaborations between libraries and other content companies, and for ProQuest and other library partners.

The project will also give the libraries experience with a big digital-publishing effort, according to Mark Sandler, a librarian

---

## Academic librarians say their involvement will help ensure that the new digital files meet the technical and scholarly standards that researchers expect.

---

at Michigan. He says Michigan was particularly interested in working with a financing model that isn't dependent on government or corporate grants, "which we don't think is sustainable in the long run."

He says that if the project succeeds in recruiting all 150 partners, the venture will have enough money to digitize 25,000 texts within about five years. Samples of the first eight texts to be digitized appear on a Michigan Web site (<http://www.lib.umich.edu/libhome/ebo>).

The keyboarding work has been subcontracted to companies in China, India, and the Philippines. Experts at Oxford's Bodleian Library are overseeing that work to ensure that the transcriptions are accurate.

"It's their cultural assets," says Mr. Sandler. He says the project could have 1,000 texts completed by the end of the year.

For the initial five years at least, only libraries that contribute to the project will be able to have free access to the new digital texts, although the goal is to make the collection more widely available to patrons of other libraries as well. Mr. Sandler acknowledges that the restriction puts the libraries in an unusual position: Typically they are the ones pressing to make materials more broadly accessible, over publishers' objections. In this case, however, he says the limits are necessary because "we have to create an incentive" for libraries to invest.

But Mr. Sandler says partner libraries

The library at Rutgers University at New Brunswick is among the early partners, urged on in part by Rudolph M. Bell, a professor of history who used Early English Books Online in an undergraduate honors seminar last year. One student used the digital images to write a paper on advice books on child-rearing of the 16th century, or "early Dr. Spock," as Mr. Bell calls it.

Others reviewed guides for midwives and pamphlets with get-rich advice, and examined frontispieces to determine the audiences for whom the books were intended.

## HELP FOR STUDENTS

Mr. Bell says having the fully digitized editions to accompany the digital images will be a huge plus. "It will be much easier for students to read through greater amounts of content," he says. There are only so many students "willing to deal with strange fonts" of the originals.

Still, he notes, the new editions will pose some challenges for students inexperienced in older forms of English. In many cases, "W" appears as "VV" in the texts, and will be keyed in that way in the new editions as well, he notes.

And as a social historian, Mr. Bell says he doesn't expect that the material he likes to use—popular books of the period—will be as high a priority for digitization as, for example, the works of John Locke.

Meanwhile, some of the libraries have begun talks with at least two other companies about similar digitizing projects. One of them has a collection of 18th-century English books; the other has microfilmed copies of early American imprints. "We would like all of these texts to integrate so we can have one huge collection," says Mr. Sandler. ■

did negotiate broad rights for use of the materials by their own professors and students, including rights to use the material in classrooms and to distribute it to students who aren't on campuses because they are enrolled in distance-education courses. "All of those uses which are a little bit contentious on our campuses," he says.

## SELLING ACCESS

ProQuest, which will co-own the new searchable edition with the library partners, will also have the right to sell access to it during the period of limited distribution, and will pay royalties to the library partners.