



For immediate release: November 12, 2008

BRIT and UNT receive more than \$700,000 to investigate digitization of herbarium collections

FORT WORTH, Texas — For 20 years, researchers have studied the dried plant specimens of Fort Worth’s Botanical Research Institute of Texas (BRIT), which houses more than one million specimens as the largest independent herbarium in the southwestern United States. Its specimens date back to the 18th century.

The Texas Center for Digital Knowledge (TxCDK) at the University of North Texas is partnering with the BRIT to develop and integrate technology that will transform data from the printed or handwritten labels of the institute’s specimens into a form that is processable by computers. BRIT and TxCDK have received a \$738,075 National Leadership Grant from the Institute of Museum and Library Services, a primary source of federal support for the nation’s libraries and museums, for the project, “High-Throughput Workflow for Computer-Assisted Human Parsing of Biological Specimen Label Data.”

BRIT and other botanical researchers rely on each plant specimen’s label — for the oldest as well as newer specimens — to provide plant species names, the names of collectors, the date the specimen was collected, and other descriptive and ecological data. However, older specimens may have labels that are difficult to decipher, at least for a computer using optical character recognition technology (OCR). In a preliminary survey of BRIT specimens, only 41 percent of the specimens’ labels could be translated into error-free, computer-readable text with off-the-shelf OCR software. The remaining 59 percent of the labels were older, poorly hand-typed or handwritten, and could not be digitized by machines alone.

Dr. William E. Moen, TxCDK director, said by creating a system where humans can work with computers to transform label data, this project will speed up the process of converting the label data into something that computers can understand “in a cost- and time-efficient manner.” BRIT plans to put the data from these labels into an online database that will be used by botanists and other scientists around the world.

“This is an essential step in helping museums and libraries to make their valuable biodiversity data available to more researchers, as well as to government agencies and others involved in conservation planning,” says Moen, the primary investigator on the project and an associate professor in UNT’s College of Information, Library Science, and Technologies, where TxCDK is housed. Moen said the research “will establish a standard model for effective conversion of specimen label data into information that will enhance the use of digital biodiversity repositories upon which so many scientists rely.”

Amanda Neill, director of the BRIT herbarium and project co-investigator, said the information on the specimen labels is critical for researchers.

“Older plant specimens may represent a final record of existence for species from habitats that are no longer intact, and may be the most valuable to researchers of global climate change, since dates of a plant’s flowering or fruiting events are recorded on the specimen,” she said. “Data from these labels can also

more

BRIT/UNT Grant
November 12, 2008
Add One

provide the most information about changes in the earth's vegetation during the last 250 years, including the movement of invasive species and the loss of endangered species over time.”

Jason Best, information technology manager for BRIT and co-investigator on the project, said digitizing biological collections in a well-planned and standardized way “increases their use by a wider audience, reduces the physical handling of the original object, and produces a permanent digital archive.”

“A high-quality digital facsimile is just as effective for research as the actual specimen in many cases, and less physical handling means less damage to these fragile and scientifically valuable items,” he said.

During the next two years, TxCDK and BRIT staff members will identify a representative sample of about 1,000 specimen labels and develop software and applications that people can use to transform the printed data. The researchers will create efficient workflows and will test and evaluate the resulting system and components to ensure the quality of the transformed data and the efficiency of the transformation process.

The National Leadership Grant, Moen said, will financially support three UNT graduate students for work on the project. Project information and results will be posted on a web site after the project officially starts Dec. 1. Research results will also be shared with managers of herbaria and members of professional societies representing other natural history collections.

For more information about BRIT, go to www.brit.org. For more information about TxCDK, contact Moen at 940-565-2473 or William.moen@unt.edu.

30

Contact:
Cleve Lancaster, Director of Development
Botanical Research Institute of Texas
817.332.4441, clancaster@brit.org