Foundations for
RESEARCH AND LEARNING

a

SHARED KNOWLEDGE. SHARED RESOURCES

There is a growing mountain of research. ... The investigator is staggered by the findings and conclusions of thousands of other workers – conclusions which he cannot find time to grasp, much less to remember, as they appear. ... Professionally, our methods of transmitting and reviewing results of research are generations old and by now are totally inadequate for their purposes.

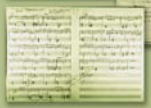
Scientist Vannevar Bush, in a 1945 *Atlantic Monthly* article proposing "Memex," a configurable device for information storage and retrieval

b) Frontispiece from "A new treatise on flower painting, or, Every lady her own drawing master," by George Brookshaw (1818). Digital Library for the Decorative Arts and Material Culture, University of Wisconsin. (http://libtext.library.wisc.edu/DLDec/Arts/)

c) World War I poster from American Women's History: A Research Guide, a Web resource developed by Ken Middleton, a librarian at Middle Tennessee State U. (http://www.mtsu.edu/~kmiddlet/history/women/wh-intro.html)

d) 19th century engraving from Tufts University's Bolles Collection on the history of London. Can be accessed through the NSF-supported Perseus Digital Library of materials on the history of Western culture. (http://www.perseus.tufts.edu)

e) March 1769 entry in Thomas Jefferson's Monticello garden book notes planting of peach, pear, cherry, apple trees. Massachusetts Historical Society digital archive is supported by Congressional Save America's Treasures program. (http://www.thomasjeffersonpapers.org)



f) Signed handwritten score of "Star Dust," from the Hoagy Carmichael Collection at Indiana University, the composer's alma mater. (http://www.dlib.indiana.edu/collections/hoagy/)

Today, more than half a century after Vannevar Bush described information overload in a data-intensive world, high-performance networking, computing, and information-management technologies are making possible the far-reaching support system for human thought that he envisioned. Although they rarely make news themselves, the diversified digital libraries of core knowledge developed through NITRD agency investments have become a necessary, and invaluable, resource for rapid innovation, not only in research but in education and training, medical practice and health care, heavy industry and manufacturing, pharmaceutical design, business-to-business technologies, agriculture, and many other fields of endeavor.

Digital collections take many forms. Some are built and managed by a single organization in one location; others are actually Web-based frameworks and protocols that enable users to access widely distributed digital archives. Materials can include images and animations, software, sound, and video as well as data and texts. In addition, NITRD-sponsored research originated the IT capabilities that make it possible for people to discover the information they need, organize it, and work with it on their desktops. These technologies – for example, Web browsers; search engines; data-mining, analysis, and management tools; metadata frameworks (information that helps users understand the origin and nature of a digital record); data display and manipulation tools; and language translation methods – provide the technical foundations for a universal digital knowledge system.

But the work is far from finished. Research across the NITRD agencies continues in FY 2003 and FY 2004 on substantial technical issues – such as interoperability among file formats, indexing protocols, and interfaces; data management, storage, and validation; networking bottlenecks; and long-term preservation – that impede development of digital libraries. These efforts reflect urgent demand in every field for deep reservoirs of sharable knowledge to maximize the value of existing findings and enhance the potential for significant advances. Some examples of NITRD digital-library achievements and works in progress:

E-PRINT Network – DOE manages the world's largest "one-stop shopping" site for preprint reports in science and technology. These papers by scientists on their unpublished research in progress are fundamental resources for the U.S. research community. The E-PRINT Alerts feature allows users to specify their interests and receive notification as relevant new information is added. Scientific papers in biology, chemistry, computer science, engineering, environmental sciences, materials science, mathematics, physics, and other areas of interest to DOE are included. (http://www.osti.gov/preprints/)

E-Specimens – NSF-funded projects at the University of Texas at Austin provide online access to libraries of biological specimens. The DigiMorph specimen library is an archive of X-ray-computed tomography data whose animations and details are used in research labs and classrooms around the world. The e-Skeletons Project lets users examine the bones of a human, gorilla, and baboon and information about them in an osteology database. (http://www.digimorph.org/ and http://www.eskeletons.org/)

NO ORDINARY TABLE

The significance of sharing resources is highlighted when the resources in question are costly specialized research facilities, such as this multi-ton "shake table" under construction at the University of Nevada at Reno as part of NSF's Network for Earthquake Engineering Simulation (NEES). Only a small number of universities have built the structurally fortified labs equipped with unique heavy machinery required to shake experimental loads on such tables with the forces of an earthquake. Yet, other than measuring damage after a real event, shake tables have been the principal scientific tool for studying how to design structures to withstand quakes.

Using high-end computing and broadband networking capabilities, the NEESgrid will enable earthquake engineering researchers nationwide to observe and participate in these experiments in real time, make use of data stored in a common archive, and share advanced modeling and simulation software to study earthquake impacts and design resilient structures.

a) Cuneiform tablets, one of humankind's earliest data storage and information management technologies (ca. 3300- 2000 B.C.), and new translation research are focus of the Cuneiform Digital Library (http://cdli.ucla.edu). Details on page 53.

g

g) Thomas Edison films in Open Video Project include first America's Cup race, Cushing/Leonard boxing match (right). (http://www.open-video.org)

Head of Herrerasaurus ischigualastensis, oldest known dinosaur, from Upper Triassic Period. Discovered in Argentina in 1959. Image from Digimorph collection, University of Texas.

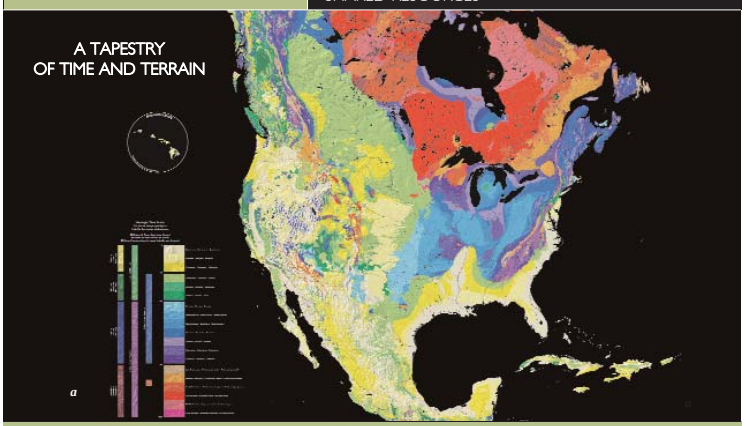
Spiders"R"Us, a University of Arizona Artificial Intelligence Laboratory program, is developing next-generation search-agent tools for knowledge discovery on the Web. (http://ai.bpa.arizona.edu/spidersrus/)

SPIDERS "R" US
Digital Library Toolkit

Foundations for RESEARCH AND LEARNING

SHARED KNOWLEDGE. SHARED RESOURCES

A TAPESTRY OF TIME AND TERRAIN



a

Digital libraries supported by the NITRD agencies are making vast stores of cultural, educational, technical, and scientific information available to the public

a) *A new online resource for students, teachers, and the public, the U.S. Geological Survey's Tapestry of Time and Terrain interactive map merges the most accurate and detailed agency data on the topography and geology of North America. Users can explore the geological history of mountain formation, river erosion and deposition, glaciation, volcanic action, and other processes over 26 billion years. Details on page 53. (http://tapestry.usgs.gov)*

GenBank and related databases – The National Library of Medicine's National Center for Biotechnology Information maintains public, searchable databases of the genomic sequences submitted in the Human Genome Project, the current draft of the genome, and completed genomic sequences for 800 other organisms. Scientists from around the world continue to upload new data to GenBank as they refine the original human draft sequences or develop sequences for other species. The Web site provides a variety of tools for working with the databases as well as archives of related information, such as a catalog of inherited human disorders. (http://www.ncbi.nlm.nih.gov/)

International Children's Digital Library – In this NSF-supported effort, the University of Maryland, the Internet Archive, and international partners are developing a library of 10,000 children's books from 100 cultures and new technologies to serve young readers. The library will serve children ages 3 to 13 worldwide. (http://www.icdlbooks.org)

NASA Cosmic Collection – In a large-scale technology transfer, the agency has released more than 500 NASA-designed software programs as publicly available open-source resources. The software, used in such fields as chemistry, aerodynamics, and engineering design, is being made available online by the OpenChannel Foundation. (http://www.openchannelfoundation.org)

National Guidelines Clearinghouse™ and National Quality Measures Clearinghouse™ – These online archives established by AHRQ provide a unique national resource for medical clinicians and health-care professionals, amassing for the first time comprehensive databases of the codified treatment knowledge, formal practice guidelines, and evidence-based quality indicators that have been developed by medical, health-care, and government organizations in clinical medicine. The guidelines clearinghouse, a partnership with the American Medical Association and the American Association of Health Plans, contains guidelines on diseases and conditions, treatments and interventions, source organizations, and related resource documents. Users can apply an online software utility to compare the contents of various guidelines. The quality measures clearinghouse provides objective summaries of quantitative measures of clinical performance (such as hospital pneumonia rates) developed by organizations throughout the health-care field. Software enabling users to compare measurement information is also provided. (http://www.guideline.gov/) and (http://www.qualitymeasures.ahrq.gov/)

National Science Digital Library – In FY 2004, NSF-funded work will continue on development of a comprehensive collection of core knowledge, teaching materials, and learning resources for nationwide high-school and college-level education in the sciences, technology, engineering, and mathematics. (http://www.nsdl.org/)

NIST Data Gateway – Through one online portal, users can access more than 80 NIST-developed databases of fundamental scientific information and validated reference data on topics ranging from atomic spectra, chemistry, mathematics, and physical constants, to calibration and manufacturing standards, product design, properties of materials, and thermophysical data. (http://srdata.nist.gov/gateway/)

Protein Data Bank (PDB) – As the international repository for 3-D data on the structures of biological macromolecules, the PDB is playing a critical supporting role in current research on the complex shapes and activities of proteins and their relationships to causes and mechanisms of disease. Funded collaboratively by DOE/SC, NIH, NIST, and NSF, this free interactive resource – accessed some 50 million times in 2002 – enables researchers to find and download validated structural data on more than 10,000 individual proteins generated with several techniques (X-ray crystallography, nuclear magnetic resonance imaging, and computational modeling). Researchers can also submit results of their own work to the collection. (http://www.pdb.org)

PubMed – The world's largest online collection of contemporary scientific literature related to the biomedical sciences and medicine, this public resource established and managed by NIH's National Library of Medicine contains 12 million abstracts from 4,500 journals. In FY 2003, it will be searched more than 400 million times. (http://www.ncbi.nlm.nih.gov/entrez/)

MEDLINEplus, a companion online library designed especially for consumers, will approach 100 million searches. (http://www.medlineplus.gov)