

Science.gov – A Single Gateway to the Deep Web Knowledge of U.S. Science Agencies

Brian A. Hitson

Associate Director, Office of Scientific and Technical Information, Office of Science,
U.S. Department of Energy

E-mail address of main author: hitsonb@osti.gov

The impact of science and technology on our daily lives is easily demonstrated. From new drug discoveries, to new and more efficient energy sources, to the incorporation of new technologies into business and industry, the productive applications of R&D are innumerable. The possibility of creating such applications depends most heavily on the availability of one resource: knowledge. Knowledge must be shared for scientific progress to occur. In the past, the ability to share knowledge electronically has been limited by the “deep web” nature of scientific databases and the lack of technology to simultaneously search disparate and decentralized information collections. U.S. science agencies invest billions of dollars each year on basic and applied research and development projects. To make the collective knowledge from this R&D more easily accessible and searchable, 12 science agencies collaborated to develop *Science.gov* – a single, searchable gateway to the deep web knowledge of U.S. science agencies. This session will describe *Science.gov* and the subsequent versions released or planned for release.

Science.gov contains reliable information resources selected by the respective agencies as their best science information. Two major types of information are included—selected authoritative science web sites (over 1700 as of early April 2004, with more being added on a regular update basis) and databases of technical reports, journal articles, conference proceedings, and other published materials. (The specific content varies by database. Thirty databases are included currently.) The selected web sites can be explored from the *Science.gov* homepage. The web pages and the databases can be searched individually or simultaneously from the search page. Version 2.0, to be released in May or June 2004, will combine the information into a single search, with results displayed in relevancy-ranked order. In all, roughly 50 million pages of R&D information are available through *Science.gov*.

Science.gov was developed by an interagency [working group](#) of 17 scientific and technical information organizations from 12 major science agencies. Together these agencies make up the *Science.gov Alliance*. These agencies are committed to serving the information needs of the science-attentive citizen, including science professionals, students and educators, business people and entrepreneurs, and members of the public with an interest in science. The Alliance and *Science.gov* were formed to improve and enhance access to information at science agency programs.

Technically, the decentralized collections of *Science.gov* are made searchable through a distributed search engine. The precision of searches has been significantly improved through relevancy ranking of title text. In the near future, precision searching will be further improved by using information in the complete bibliographic record and abstracts, when available. Plans are also being made for *Science.gov* to take precision searching to the next level, basing its relevancy ranking on full text.