



Date : 23/06/2008

WorldWideScience.org – the Global Science Gateway
Bringing the World's Science to all Corners of the Globe

Walter L. Warnick, Ph. D.
Director, U.S. Department of Energy Office of Scientific and Technical Information (Operating Agent for WorldWideScience.org)
Washington, D.C., United States

Meeting: 109. Science and Technology Libraries
Simultaneous Interpretation: English, Arabic, Chinese, French, German, Russian and Spanish

WORLD LIBRARY AND INFORMATION CONGRESS: 74TH IFLA GENERAL CONFERENCE AND COUNCIL
10-14 August 2008, Québec, Canada
<http://www.ifla.org/iv/ifla74/index.htm>

WorldWideScience.org
The Global Science Gateway

Walter L. Warnick, Ph.D.
Director
United States Department of Energy
Office of Scientific & Technical Information
(Operating Agent for WorldWideScience.org)
Washington, D.C., United States



Abstract

WorldWideScience.org, a global science gateway, explores new trends in global scientific communication, including how technology opens borders by providing access both to sources and to users located in diverse geographic settings. International collaboration is a key component in WorldWideScience.org's success in

providing scientific information to a global population. WorldWideScience.org was designed to accelerate scientific discovery and progress by accelerating the sharing of scientific knowledge. Through a multilateral partnership, WorldWideScience.org enables anyone with internet access to launch a single-query search of over 30 national scientific databases and portals in many countries. From a user's perspective, WorldWideScience.org makes the databases act as if they were a unified whole. The revolutionary technology behind WorldWideScience.org will be discussed.

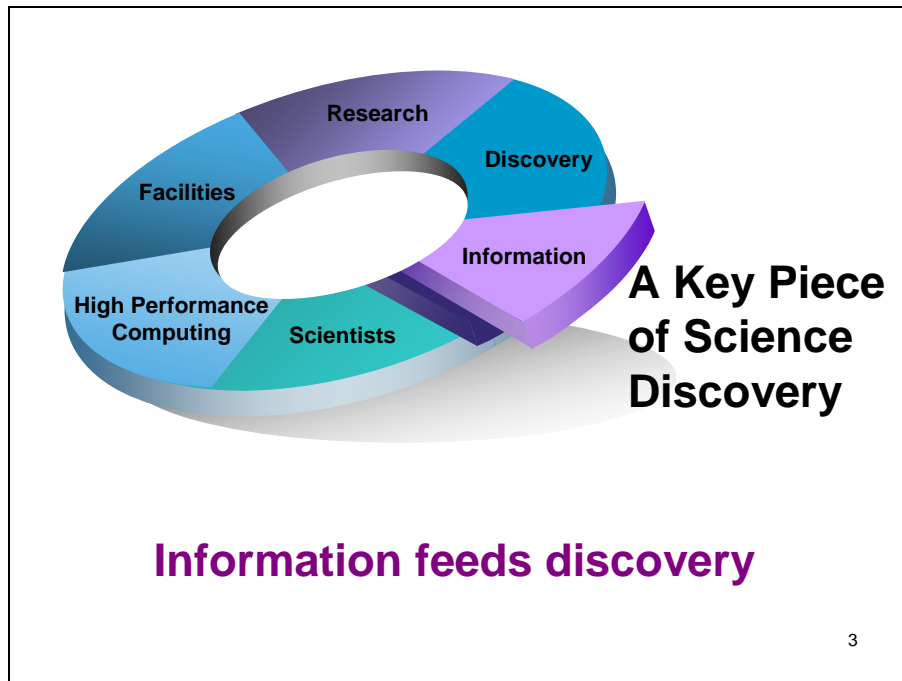
Science Progresses as Knowledge is Shared

**If the sharing of knowledge is
accelerated, discovery is
accelerated**

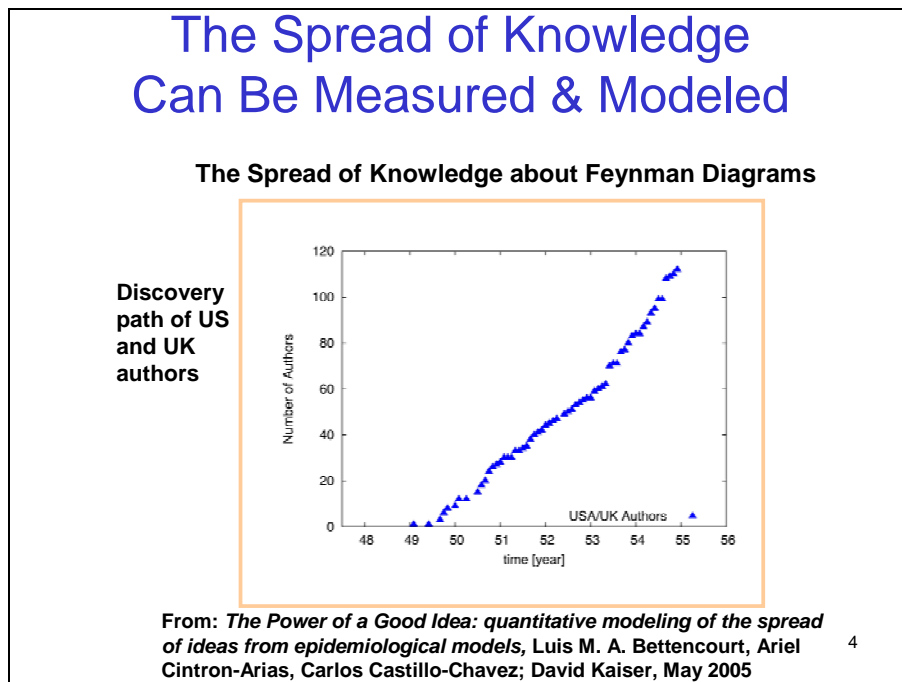
Profound implications for **all of us**
in the information business!

2

WorldWideScience.org is a global science gateway designed to accelerate scientific discovery and progress by accelerating the sharing of scientific knowledge. Every scientist, regardless of discipline, will agree that science progresses only if knowledge is shared among colleagues.

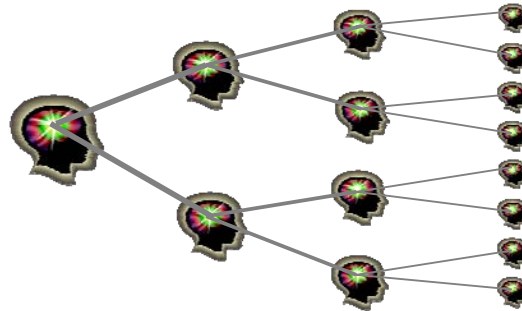


Thus, if the sharing of knowledge is accelerated, discovery is accelerated. Discovery powers the growth of world prosperity, discovery improves people's lives, and discovery accelerates science.



The spread of knowledge can actually be measured. Studies have indicated that knowledge spreads in much the same way as diseases. Of course, the spread of knowledge is a more desirable and appealing phenomena to study. Epidemiological diffusion models created to track the spread of disease can also be applied to track the spread of knowledge.

The Spread of Knowledge Can Be Accelerated



Knowledge is contagious.

**Increasing the contact rate means
researchers “catch” an idea faster.**

5

A key parameter of the model is the “contact rate.” By increasing the contact rate, the spread of knowledge is accelerated. To that end, products like WorldWideScience.org help researchers “catch” ideas faster.



International partnership kicks off global science gateway



In January 2007, Dr. Raymond Orbach, DOE Under Secretary for Science, and Lynne Brindley, Chief Executive of the British Library signed a Statement of Intent to partner in the development of a searchable global science gateway.

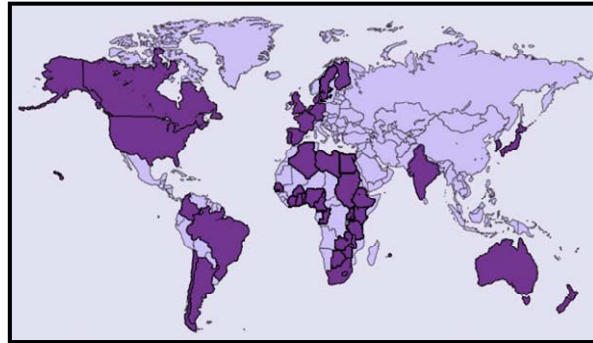
6

First envisioned by a member of the International Council for Scientific and Technical Information (ICSTI) at the June 2006 meeting in Bethesda, Maryland, U.S., the concept was strongly endorsed by the British Library, which offered to collaborate with the United States Department of Energy (DOE). A bilateral partnership was

formalized in January 2007 when the British Library and DOE signed a Statement of Intent to partner in the development of a Global Science Gateway. Dr. Raymond Orbach, DOE Under Secretary for Science, and Lynne Brindley, Chief Executive of the British Library, participated in the signing ceremony at the British Library. The Statement of Intent invited other nations to join the collaboration.

WorldWideScience.org was launched in June 2007

Now searches 32 portals from 44 countries



7

Later officially named “WorldWideScience.org,” the gateway was developed by the U.S. Department of Energy’s Office of Scientific and Technical Information and was unveiled to ICSTI members and the public at the June 2007 ICSTI meeting in Nancy, France. ICSTI members also endorsed the proposal to serve as the umbrella organization for the future long-term governance of WorldWideScience.org. A Terms of Reference document was developed to define this governance structure, and the Terms of Reference were accepted by ICSTI members at the 2008 Winter meeting in Paris.



- **A federation of the leading science portals sponsored by the governments and national institutions of 44 countries**
- **A quantity of science (more than 200 million pages from every inhabited continent)**
- **A breakthrough in content enabled by breakthrough technology**

8

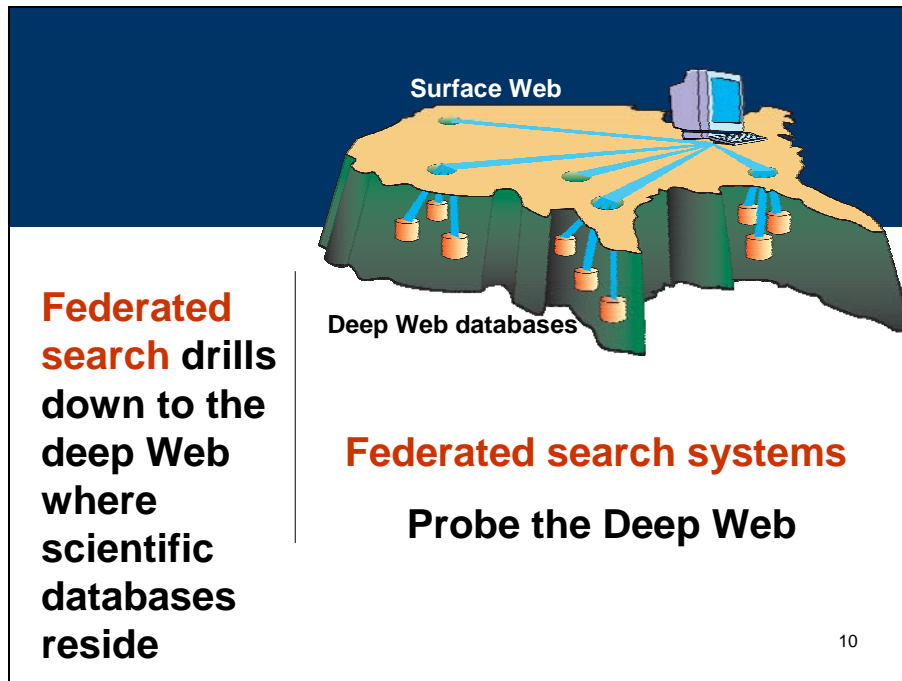
WorldWideScience.org enables anyone with internet access to launch a single query search of over 30 national scientific databases and portals in more than 40 countries, covering all of the world's inhabited continents and nearly half the world's population. Many of the geographic areas included in WorldWideScience.org, such as some of the African countries included in African Journals Online, are not traditionally well represented in standard scientific and technical resources. WorldWideScience.org and its unique technology improve the visibility of R&D outputs and expand global scientific communication by providing equal access to diverse geographic settings.

WorldWideScience.org & Federated Search Technology

Many popular search engines rely on crawler based technology

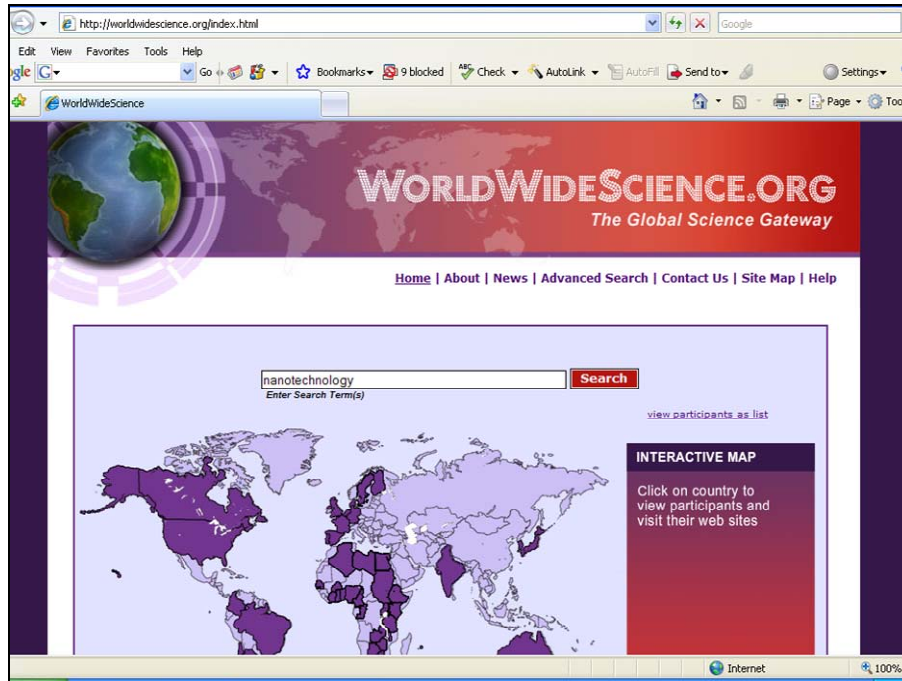


WorldWideScience.org implements federated searching to provide its encompassing coverage of global science and research results. It is a little known fact that many of the popular search engines overlook a large portion of the web. Their technology relies upon crawlers, which find and visit web sites one at a time by following hyperlinks. Each time a crawler finds a page, it indexes it. The index is then merged with the master index, and when the user does a search, the query is actually applied against the master index. When there is a match, the results are to hyperlinks indexed sometime in the past.

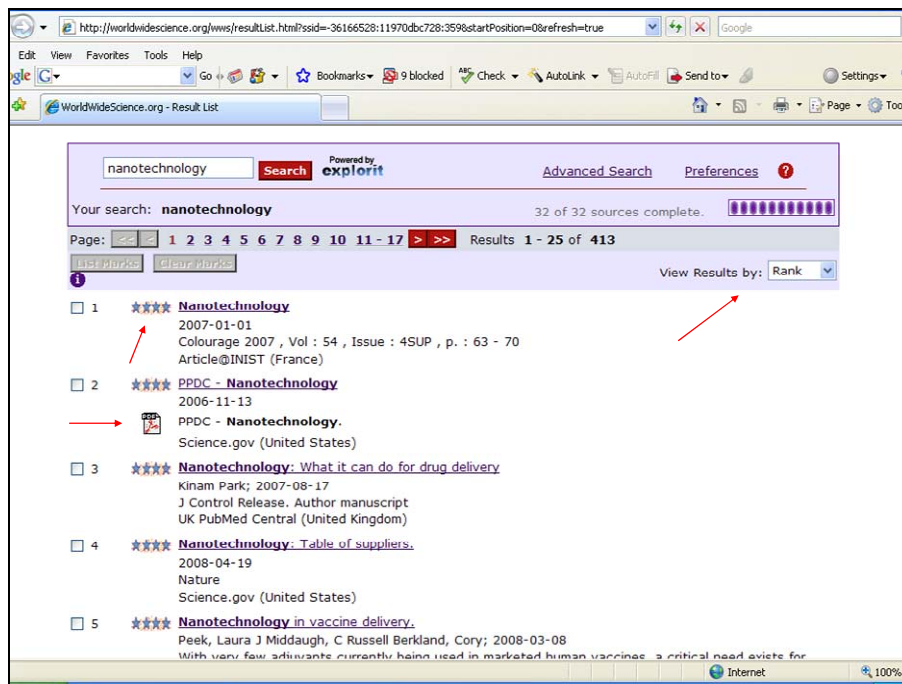


The bulk of science information, especially scholarly science information, resides in databases. Crawlers can get to the first page of a database, but, typically, they cannot get past the front page. The database's search box is often the only systematic way to see the contents of the database, and crawlers are unable to process the search box. This part of the web that is off limits to crawlers is called the Deep Web. It is possible for database owners to take special steps to expose their database content to crawlers; however, many organizations have resource constraints and do not pursue these options.

Federated search is a different kind of web search architecture. When the user places a query on a federated search application, like WorldWideScience.org, the query is transmitted to all the servers that host the databases. Those servers then translate the query into its own database and execute the search. Each remote database reports its results back to the WorldWideScience.org server, which combines the hits from all the databases, and sorts them in relevance ranked order. Finally, the ranked list is returned to the user. The whole process can take anywhere from about a second to around twenty seconds, depending on the complexity of the search and the speed of the source databases. Thus, WorldWideScience.org allows the user to search multiple data sources with a single query in real time.

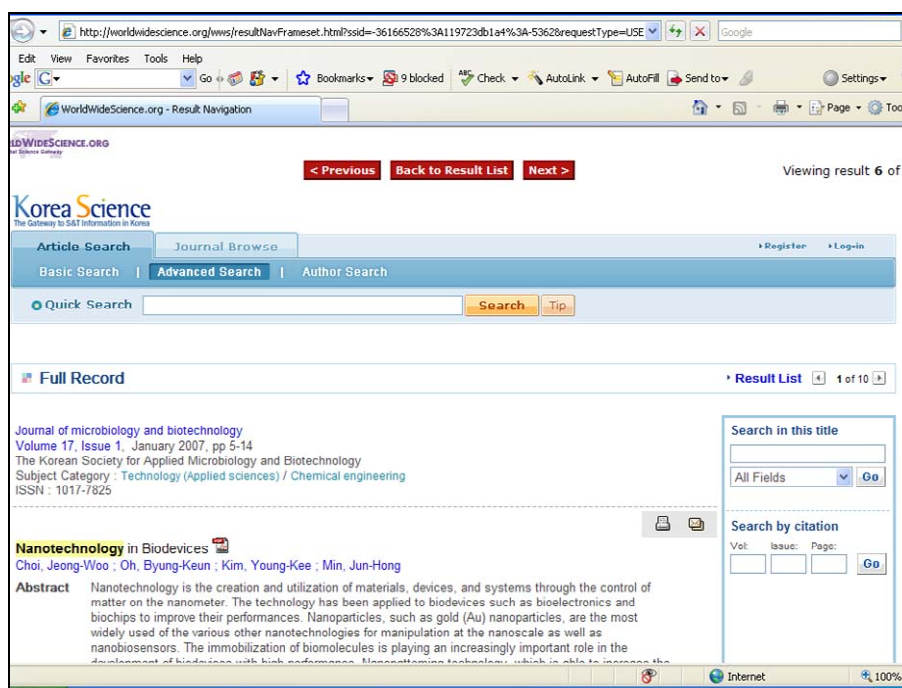


Providing a specific example, if a user searches on the term “nanotechnology”, the WorldWideScience.org interface sends the query to the 32 source databases, which independently run this search and begin returning results.

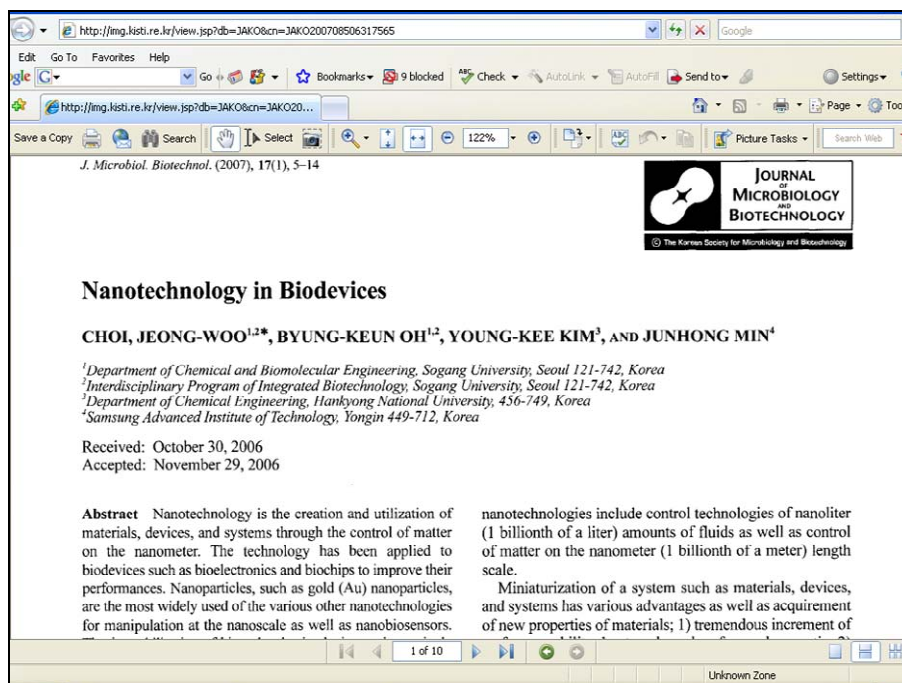


As results are returned to WorldWideScience.org, the combined results from all sources are run against WorldWideScience.org’s relevance-ranking algorithm and presented to the user based on the prevalence of the search term in the title, metadata,

and any other snipped information provided by the source. The user also has the flexibility to reorder the results by source, date, title, or author.



Users can view the complete record for each result, and the full-text document if it is available.



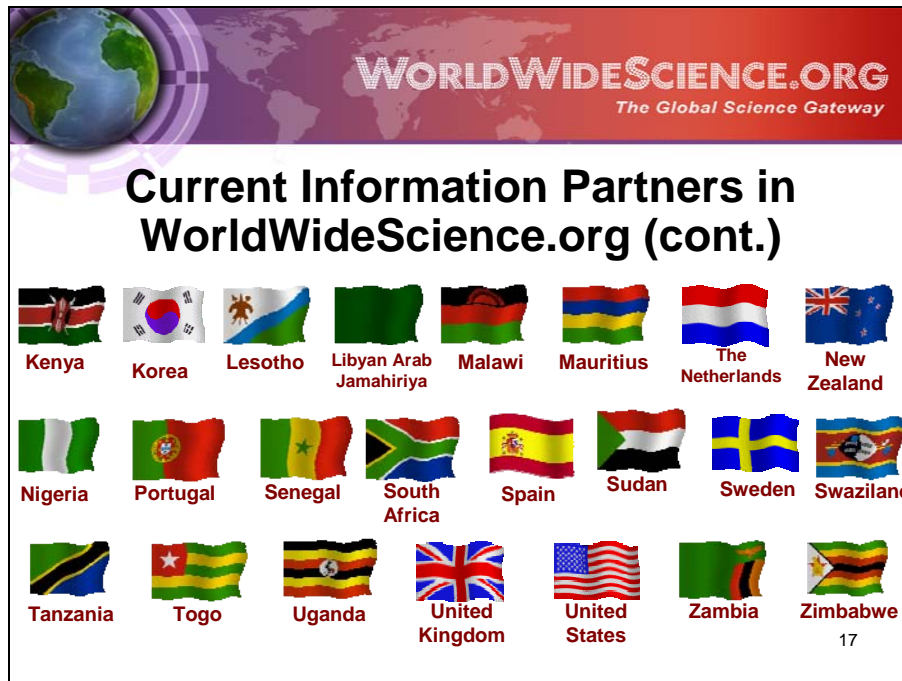
With a large number of open access sources, WorldWideScience.org provides a single point of access to vast quantities of full-text science literature.

Current WorldWideScience.org Sources

- African Journals Online
- Article@INIST (France)
- Australian Antarctic Data Centre
- Canada Institute for Scientific and Technical Information
- CSIR Research Space (South Africa)
- Defence Research and Development Canada (Canada)
- DEFF Global E Prints (Denmark)
- DEFF Research Database (Denmark)
- Directory of Open Access Journals (Sweden)
- Electronic Table of Contents (ETOC) (United Kingdom)
- Indian Academy of Sciences
- Indian Institute of Science Eprints
- Indian Institute of Science Theses & Dissertations
- Indian Medlars Centre
- J-EAST (Japan)
- J-STAGE (Japan)
- J-STORE (Japan)
- Journal@rchive (Japan)
- Korea Science (Korea)
- NARCIS (Netherlands)
- Science.gov (United States)
- Scientific Electronic Library Online (Argentina, Brazil, Chile, Colombia, Portugal, Spain)
- Transactions and Proceedings of the Royal Society of New Zealand 1868-1961 (New Zealand)
- UK PubMed Central (United Kingdom)
- Vascoda (Germany)
- VTT Technical Research Centre of Finland Publications Register
- VTT Technical Research Centre of Finland Research Register

Current Information Partners in WorldWideScience.org





Since its release in June 2007, WorldWideScience.org has more than doubled the number of data sources searched, along with greatly increasing the number of countries participating as information providers. It provides access to large prominent collections such as Science.gov (the U.S. contribution) in addition to less well-known sources of highly valuable science. It is estimated that WorldWideScience.org covers more than 200 million pages of scholarly scientific content.

Founding Alliance Members

- Canada Institute for Scientific and Technical Information (CISTI) – Canada
- VTT Technical Research Centre of Finland (VTT) – Finland
- Institut de l'Information Scientifique et Technique (INIST) – France
- Korea Institute of Science and Technology Information (KISTI) – Korea
- Scientific Electronic Library Online (SciELO) – Argentina, Brazil, Chile, Colombia, Portugal, Spain
- Council for Scientific and Industrial Research (CSIR) – South Africa
- British Library – United Kingdom
- Science.gov Alliance – United States
- International Council for Scientific and Technical Information (ICSTI)

18

Along with vastly increasing its content since its inception, WorldWideScience.org has also transitioned from bilateral management to a multilateral governance structure, called the WorldWideScience Alliance. The Alliance consists of eight

founding member organizations representing 13 countries (Argentina, Brazil, Canada, Chile, Colombia, Finland, France, Portugal, South Africa, South Korea, Spain, United Kingdom, and United States). In addition to member countries, ICSTI also serves as an Alliance member and primary sponsor.



Alliance Executive Board:

- Chair – Richard Boulderstone, British Library
- Deputy Chair – Pam Bjornson, Canada Institute for Scientific and Technical Information
- Treasurer – Tae-Sul Seo, Korea Institute of Science and Technology Information
- Ex-Officio Member – Walter Warnick, WorldWideScience.org Operating Agent, U.S. DOE Office of Scientific and Technical Information
- Ex-Officio Member – Herbert Gruttemeier, ICSTI President, French Institut de l'Information Scientifique et Technique
- At-Large Member – To be Determined

19

An election for the Alliance's Executive Board was held in early April 2008. Richard Boulderstone, Director, E-Strategy and Information Systems, British Library, was elected Chair. Pam Bjornson, Director General, Canada Institute for Scientific and Technical Information, was elected Deputy Chair. Tae-Sul Seo, Senior Researcher, Korea Institute of Science and Technology Information, was elected Treasurer. The Executive Board also includes two ex-officio members, WorldWideScience.org Operating Agent Walter Warnick, Office of Scientific and Technical Information, U.S. Department of Energy, and ICSTI President Herbert Gruttemeier, French Institut de l'Information Scientifique et Technique. An "at-large" delegate from the Alliance will also be appointed to the Executive Board.

A formal ceremony commemorating the establishment of the Alliance was held in Seoul, Korea on June 12, 2008. Founding members of the Alliance participated in the events. The Alliance welcomes new members and is particularly interested in engaging the participation of main science-producing nations, such as Russia and China.

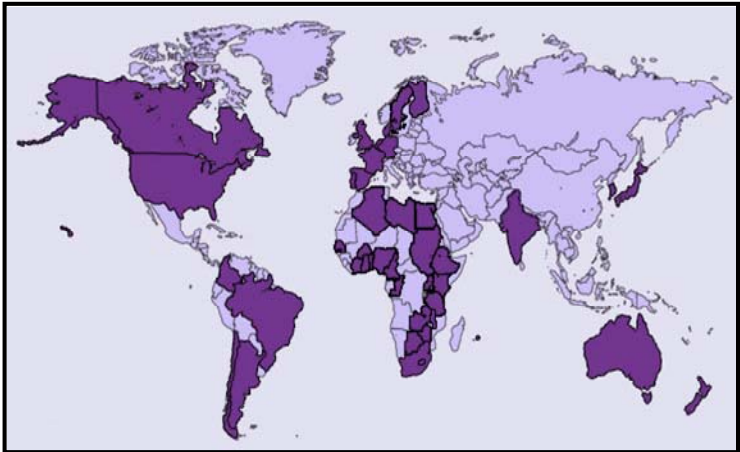



Recent & Future Enhancements:

- Results Clustering
- Personalized Alert Service
- Translation Capabilities

20

Along with increasing the number and diversity of scientific sources searched by WorldWideScience.org, the Alliance has several other near-term goals. Web 2.0 functionality, such as Alert services and clustering of results, will be added later in 2008. The Alert service will allow users to set up a profile and then generate automatic queries against the WorldWideScience.org sources on a routine basis. So, a user that is interested in seeing all new documents on a particular subject can have those results delivered to his email account weekly. Clustering of results based on similar keywords and concepts will also add value for the user. The integration of translations tools is also being considered for later in 2008 and 2009.



21

Through such efforts, WorldWideScience.org is well timed to other trends in global scientific communication. National research organizations recognize the importance of increasing visibility of their R&D outputs, even in very small countries. At the same time, full-text information accessibility has increased. Through the innovative combinations of federated search and other technologies, scientists and citizens across the globe now have unprecedented access to scientific knowledge.