

Science Research: Journey to 10,000 Sources



Presented by:
Abe Lederman, President and Founder
Deep Web Technologies, Inc.

Special Libraries Association Annual Conference – June 15, 2009

Global Discovery on the Internet: A Grand Challenge

Symposium: AAAS Annual conference
in February, 2006

Three Important Observations:

1. **Scientific progress depends on the diffusion of knowledge;**
2. **Knowledge that may lead to breakthroughs frequently resides in distant scientific communities; and**
3. **Innovation is needed to speed up the diffusion of knowledge.**

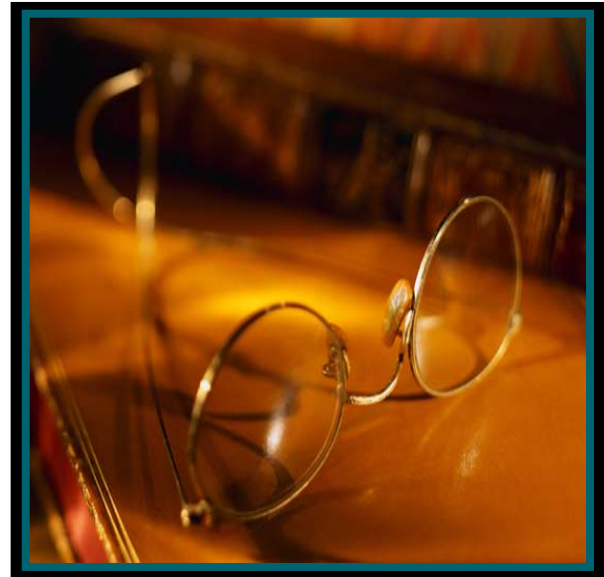
Mapping the Journey



- The Vision
- Federated Search
- The Challenges
- Where We Are
- Where We're Going

The Vision

- Help to advance scientific progress
- Connect people in distant scientific communities
- Build a science search portal that queries 10,000 sites at once



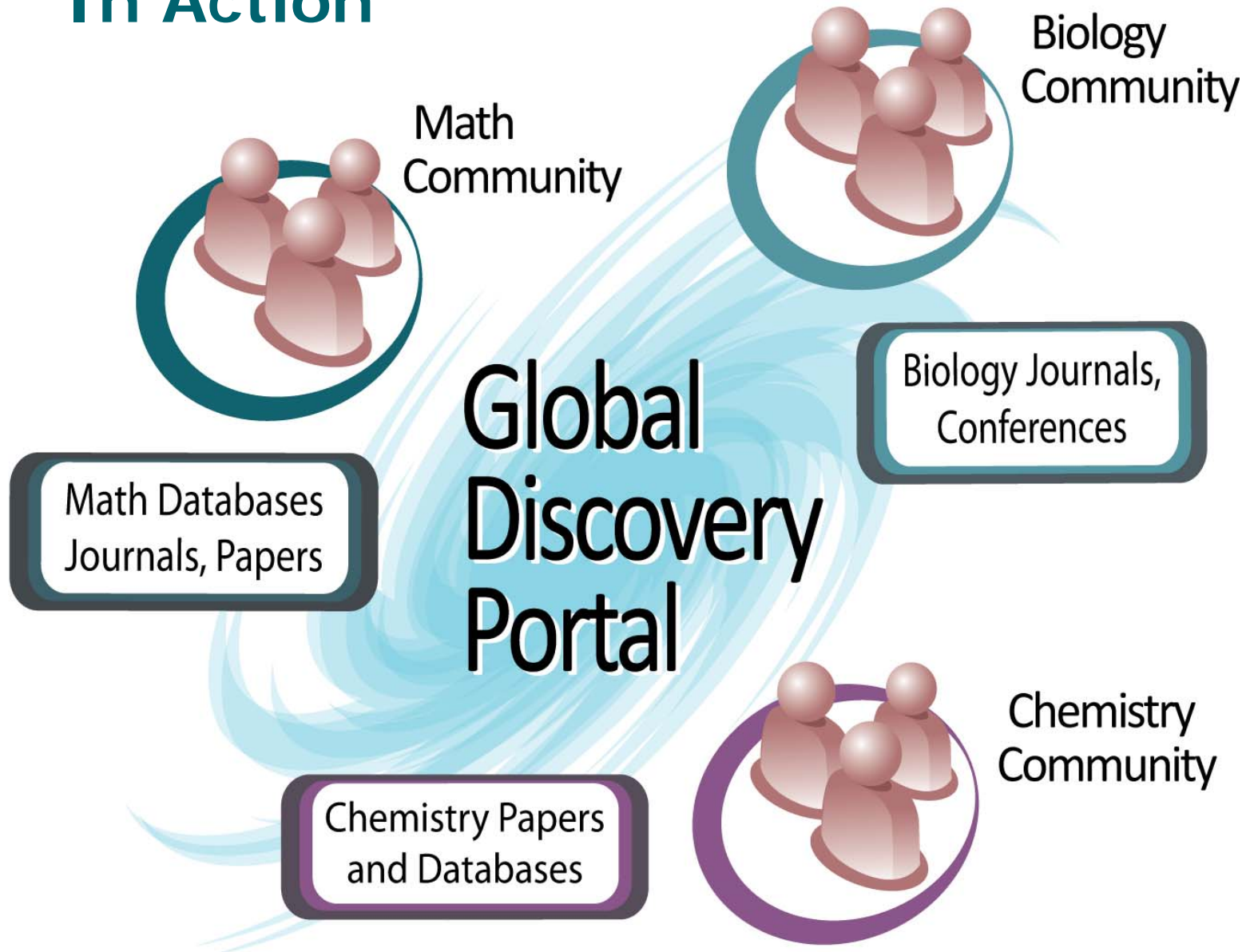
Global Discovery has the same goal as...



"...the fabled Library of Alexandria, namely to make all of science available in one place. Except in this case the place is everywhere at once, because anyone in the world could access the Global Discovery facility."

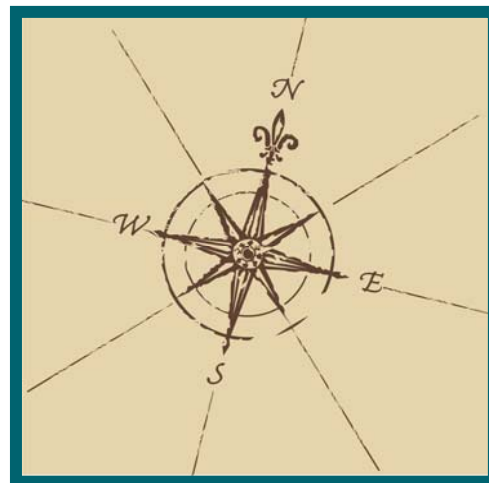
Dr. Walter Warnick, Director of the U.S. Department of Energy Office of Scientific and Technical Information (OSTI).

Knowledge Diffusion In Action



What Is Federated Search?

Federated Search is an application or service that allows a user to submit a search in parallel to multiple, distributed information sources and retrieve aggregated, ranked and de-duped results.



Federated Search Supports the Vision

- Facilitates cross-fertilization of ideas by searching diverse sources
- Helps accelerate discovery by providing access to only quality, not easily accessible content
- Allows research output to quickly be made available around the globe



The Challenges

- Source selection
- Ranking and organizing of results
- Traffic management
- System load management
- Finding, building, and maintaining connectors



Source Selection

- Match user queries to the best available sources automatically
- Prevent overwhelm of having to select from among thousands of sources
- Profile sources
- Maintain history of sources that returned best results

Ranking and Organizing of Results

Solar Power Satellite: Direct Current (DC) Power Distribution
★★★★★
2000-01-01
American Society of Mechanical Engineers

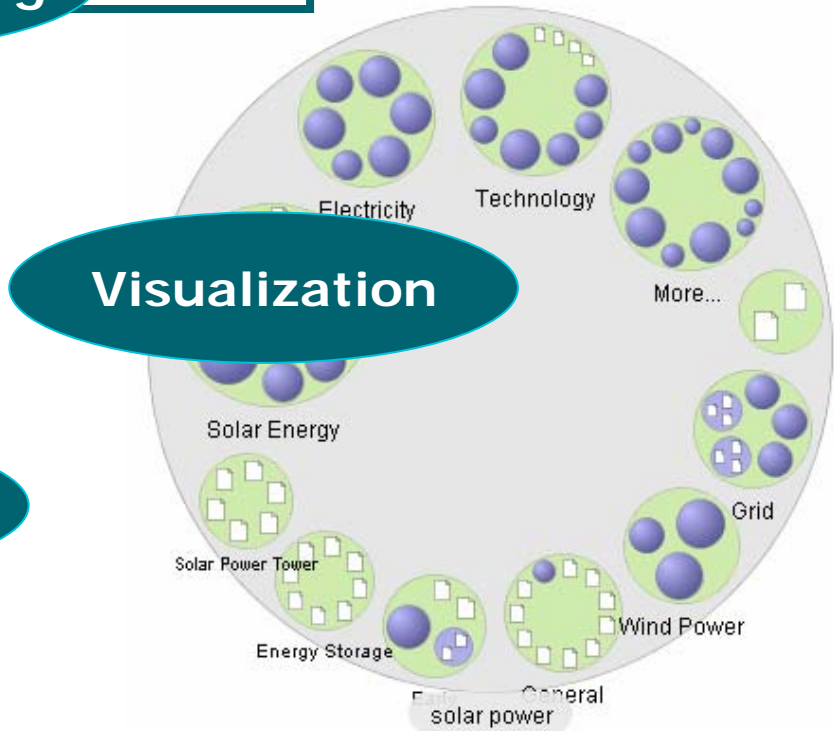
Relevance Ranking

● **All Results (4010)**

▼ **Topics**

- ▶ Solar Energy (402)
- ▼ Distribution System (338)
 - ▼ Generation (86)
 - Wind (86)
 - More...
 - ▶ Photovoltaic (87)

Clustering



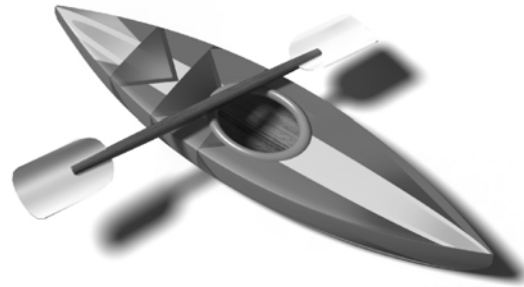
Traffic Management

- Goal is to minimize number of queries to sources
- Cache queries and their results
- Provide canned results where appropriate
- Throttle search requests to content providers during their peak load



System Load Management

- Need to be able to scale to handle more workload
- Distribute computation among "nodes"
- Use divide-and-conquer approach
- "Cloudify"



Finding, Developing and Maintaining Connectors

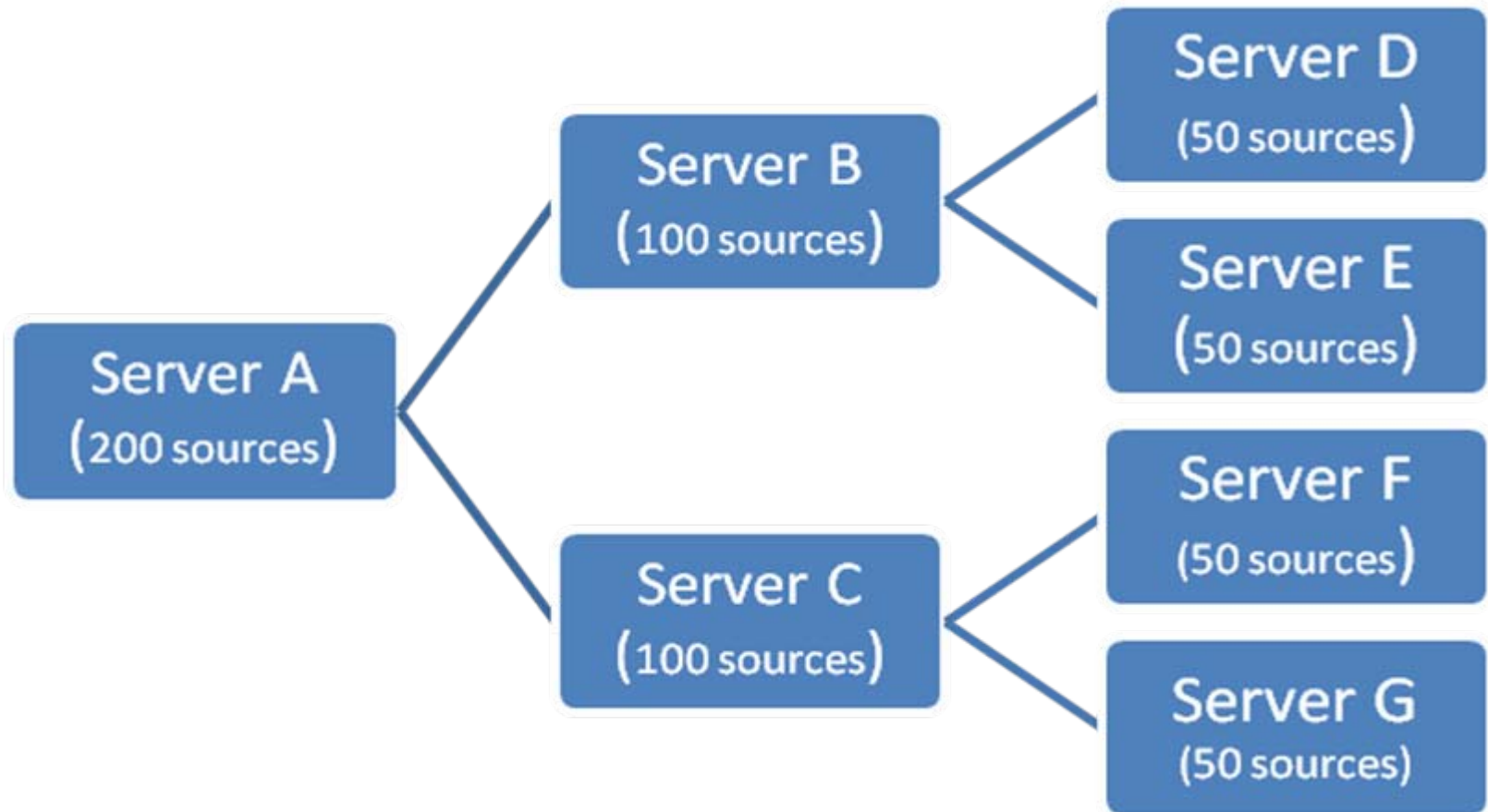
- Develop tools that mine the web for new sources
- Improve the tools and processes to build, test, monitor, and repair connectors
- Document the process of creating connectors, then...
- Crowd-source their creation and maintenance

Where We Are Now



- Divide-and-Conquer enables scalability
- WorldWideScience.org brings together the global science community
- ScienceResearch.com advances scalable federated search

Divide and Conquer



WorldWideScience.org



WorldWideScience.org
launched in June 2007

Enter Search Term(s) [Search](#)

[view participants as list](#)

UNITED STATES

U.S. Departments of Agriculture, Commerce, Defense, Education, Energy, Health and Human Services, and Interior; Environmental Protection Agency, Library of Congress, National Aeronautics and Space Administration, National Science Foundation, United States Government Printing Office, and National Archives and Records Administration.

- [Science.gov](#)

ScienceResearch.com

[Home](#) | [About](#) | [Contact Us](#) | [Help](#)



Science
Research.com
beta

The World's Science All in One Place

Search

[Advanced Search](#)

[Preferences](#)



Quick Links:

[About](#)

[FAQ](#)

[Suggest a Collection](#)

[Help](#)





[Press Room](#)

[Release Notes](#)

[Get your Own ScienceResearch.com](#)

ScienceResearch.com
Launched
June 15, 2009

Advanced Searching

Featured Collections:  [science.gov](#)  [Oalister](#)  [WorldWide Science.Org](#)  [E-Print-Network](#)

Patents Science News

Full Text:

Title:


Author:

Match:

Date Range: to

- Agricultural Sciences
- Astronomy & Space
- Biology & Nature
- Chemistry
- Computer Science
- Defense
- Earth & Planetary Sciences
- Energy
- Health Sciences
- Materials
- Mathematics
- Multidisciplinary
- Physics

Chemistry

Featured Collections:  [ACS Publications](#)

Select All

[Annual Reviews](#)

[ChemID Plus](#)

[Directory of Open Access Journals](#)

[Electrochemical Society](#)

[HighWire Press](#)

[Hindawi Publishing Corporation](#)

[IngentaConnect](#)

[Intute](#)

[IUPAC](#)

[J-Stage](#)

[National Academies Press](#)

[National Technical Information Service](#)

[Nature Publishing Group](#)

[NIST Chemistry Web Book](#)

[Oxford University Press](#)

Full Text:

Title:

Author:

Match:

Date Range: to

ScienceResearch.com would like to thank Grace Baysinger, for volunteering to be the editor for our Chemistry collection.

Grace Baysinger is the Head Librarian & Bibliographer, Swain Library of Chemistry and Chemical Engineering, Stanford University.

Please contact [Grace Baysinger](#) for questions or collection suggestions.

Your search: **Full Text: renewable energy** yielded 5,271 top results from at least 92,860,824 found.

renewable energy

New Search

[Advanced Search](#)

435 of 435 collections complete

TOPICS

[All Results \(5291\)](#)

▼ Topics

[Renewable Energy](#)

Sources (384)

[Energy Systems \(299\)](#)

[More...](#)

▼ Authors

[None \(21\)](#)

[Corkery, Pat \(19\)](#)

[More...](#)

▼ Publications

[Soil Sci Soc Am J \(40\)](#)

[ASME Conf. Proc.](#)

Results 1 – 10 of 5,291 Sort by: Rank



Limit to: All Collections



1 [renewable_energy](#)

★★★★★

Read the **Energy** in Brief on **Renewable Energy**
[Energy Information Administration](#)



2 [Renewable energy](#)

★★★★★

This resource follows the **renewable energy** thread from the Building **Energy** Efficiency Research (BEER) project at the University of Hong Kong's Architecture Department. There is a series of introductory lecture notes on the subject of **renewable energy**, accompanied by a list of case studies and related resources. This includes international examples of **renewable energy** projects, as well as profiles and contact details of appropriate institutions working on **renewable energy** research.
[Intute](#)



3 [Renewable energy](#)

★★★★★ *Rohn, Christine R.*

GREAT LAKES SEAWAY REVIEW, Vol. 35 No. 4 2007-01-01
[Science.gov \(United States\)](#)

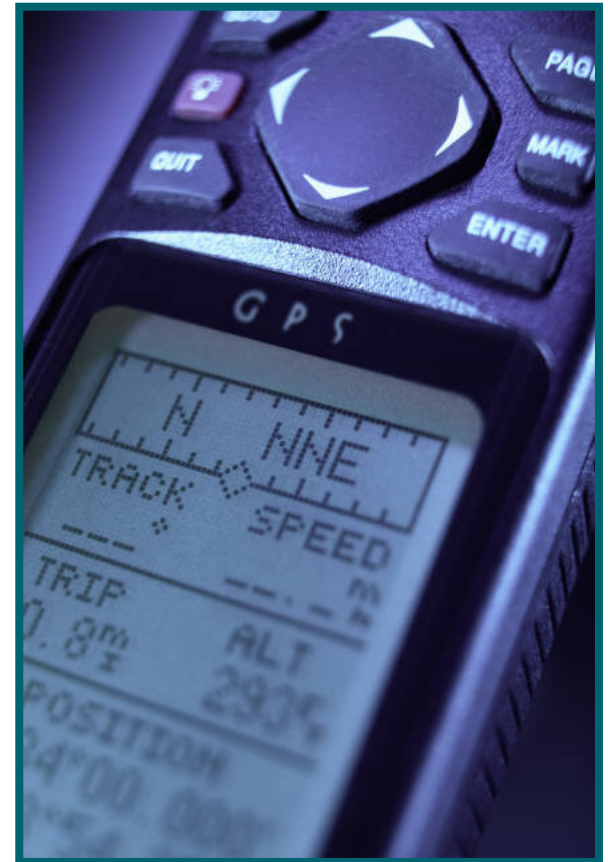


4 [Renewable Energy](#)

★★★★★ *Boyle, Godfrey*

Where We Are Going

- Search sources in any language
- Build social network to facilitate global dialog about the portal and its sources
- Build a number of science search portals and combine them into a portal with 10,000 sources



Thank You!



Contact me via email:
abe@deepwebtech.com

