

PARTNERING IN SCIENCE INFORMATION: NECESSITIES OF CHANGE

A public conference organized by the International Council for Scientific and Technical Information ([ICSTI](#)) and held at the National Library of Medicine ([NLM](#)), Bethesda, Maryland, USA, June 7-8, 2006.

A report by [Wendy Warr](#)

DAY ONE. THE NEW PLAYERS AND RELATIONSHIPS

Welcome and Introduction

[Bernard Dumouchel](#) of the Canada Institute for Scientific and Technical Information ([CISTI](#)) gave a welcoming address outlining the unique position of [ICSTI](#) since its members are a broad church, representing scientific users, publishers, national scientific and technical information (STI) centers, international unions and so on. He thanked [CSA](#) [the abstracting and indexing company, not the Chemical Structure Association Trust] and the US Geological Survey National Biological Information Infrastructure (USGS [NBII](#)) for hosting the meeting and [NLM](#) for the use of the Lister Hill Auditorium. [NERAC](#) and the [Library of Congress](#) also sponsored the meeting.

[Don Lindberg](#), Director of [NLM](#), made some opening remarks. There has been a revolution in the way research results are published and shared. The National Institutes of Health (NIH) have stated that research that they fund should be made publicly available in PubMed Central but so far this has been a request, not a mandate. Lindberg thinks that Congress will mandate it fairly soon. Only 10-15% of currently published papers directly credit NIH research support, and hence would be required to submit full text to PubMedCentral. The Howard Hughes Institute and the Wellcome Foundation have mandated open access to the results of research they fund, as have the Max Planck Institutes, Centre National de Recherche Scientifique (CNRS), Deutsche Forschung and others.

But what about raw data? A milestone was the human genome project. It finished early and under budget. It was a major success with participants in many nations. Forty-two per cent of data in the human genome project were from outside the United States, so it is fair that 40% of (free) access to [NLM](#) databases is from outside the United States. [NLM](#) also makes a single nucleotide polymorphisms (SNP) database freely available. A distributed international program gathered those data.

Clinical trial data introduce a new proposition. Congress was becoming frustrated because their constituents did not have access to trial data, so [NLM](#), NIH and the Food and Drug Administration (FDA) were told to produce a database. After two years, data from 7,000 trials had been collected but no-one knew how many trials could be included. Perhaps one could apply the *Journal of the American Medical Association* criteria: "Is it new, is it true, is it important?" The public was concerned that pharmaceutical companies were withholding negative trial data (e.g., about adolescent suicides) and that people were volunteering for trials but were not being informed about trial results. The *Annals of Internal Medicine*, the *Journal of the American Medical Association* and the *New England Journal of Medicine* agreed that there had to be a change. There are now 29,500 trials in the [NLM](#) database. People want these trials to be indexed in Index Medicus.

[NLM](#) databases attract more than one million queries a day even if the average man in the street does not know about the good work that [NLM](#) does. At least six full-time equivalents are dedicated to database security. Many, many people try to hack into the system. That's one lesson that [NLM](#) has learned but the three major lessons are as follows. Broad access to digital data is essential to biomedical research. Curation of databases is essential for their proper use. An extensive infrastructure is needed to develop and maintain the system.

Keynote Talk

World Upside Down: the Era of the Individual

[Leigh Watson Healy](#), [Outsell, Inc.](#)

A salutary tale of a company that has re-invented itself is that of Kodak. Its motto used to be “you press the button, we do the rest”. Kodak does not make film cameras any more. Getting hold of film will soon be impossible. Canon and Olympus are also re-inventing themselves. Healy showed diagrams of information industry relationships and distribution channels in the publishing industry pre-1990, in the 1990s, and after the dot.com collapse. The diagram has become progressively more complex: it has morphed into what Healy calls “content spaghetti”. Today it is the individual who is in control. Healy drew a parallel between the Kodak story and the information industry.

Google is a super power with a war chest of tools. [Outsell](#) estimates that information is a \$384 billion market (excluding government databases) the growth of which is slowing. The growth pattern mirrors that of the US economy but is six months ahead. STM is only 5% of the information “pie”; news takes up 27%. In 2005, \$3 billion of revenues were added by the top 10 information companies (Reed Elsevier, Thomson, Time Warner, Gannett, Pearson, McGraw Hill, Reuters, VNU, The Tribune and News Corporation) while Google and Yahoo increased revenues by \$4.7 billion. [Outsell](#) projects that the total industry will grow by 8.0% over the next few years, led by growth in search. STM growth of 5.8% is predicted. [Everything in this paragraph is proprietary [Outsell](#) analysis.]

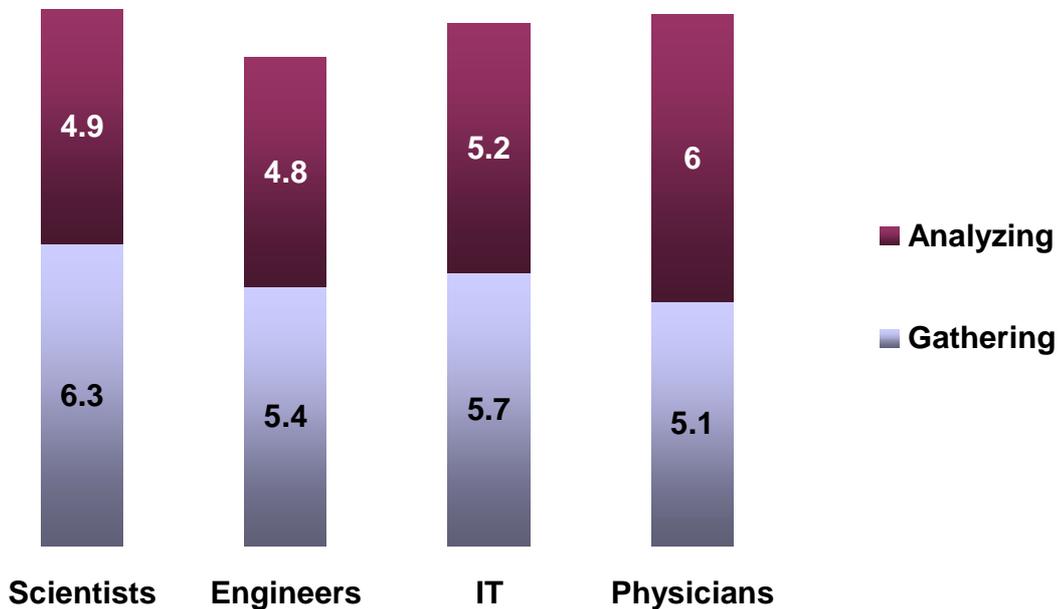
Disruptive technologies have appeared in cycles: CD-ROM, then the Web, then XML, then Web services, then Web 2.0, then AJAX. In 2006 we are in the era of Web 2.0: the Web of relationships, the Web of empowerment, and the Web as a platform. This frees us from the restrictions of software on the hard drive. Content will no longer stay in containers; packaging it is on the way out. Even selling one chapter of a book rather than the whole book is not going far enough.

Content *looks* to the user as if it were free. It is not free (the advertiser is paying) but the user thinks that it is free. This commoditization effect is challenging for the content producer. Like products provide alternatives for the user. There are two customer types: the expert and the beginner. A customer focus emerges; partners become competitors; competitors become partners; and segments and business models fall apart. There is a power shift towards the individual. The data-to-wisdom pyramid is turned upside down. The “peer information down” pyramid has a market of one at the bottom. Nowadays, eBay lets us know the price of everything; in the past no-one knew what price an item might command. There has been a move from product-centered to market-centered.

The information user of today is a “young digital native”. We have seen the rise of amateur professionals with blogs etc. Users have shrinking attention spans. Kids are “polychronic”: they can be in multiple conversations at once in chat rooms and on cell phones. There is a reliance on peer networks. These digital natives are entering the workforce and baby boomers are exiting. Content is commoditized; the Internet is like oxygen or electricity.

Healy showed a photograph of a teenager with ear buds in her ears, a camera on her chest, sending emails and text messages at the same time. Two kids will share a pair of ear buds, using one ear each to listen to a shared iPod. This is the era of information as entertainment and entertainment as information. We have to capture the minds of these digital natives.

It is disturbing how much loss of productivity is shown by the figures [Outsell](#) has found for hours per week spent using information:



Source: Outsell's Information Markets and Users Database
 © 2006 Outsell, Inc. All rights reserved.

The situation has been getting worse over the last four years. Physicians are the only group who spend more time analyzing information than gathering it. About 70% of scientists go to the Internet first when they are looking for information. The proportion of scientists who turn to the intranet for their next option is 20%. Physicians use the library, rather than the intranet as their second resort. Physicians are noted for putting in only one word for an Internet search. There has been a tremendous increase in the number of vertical search portals to address this problem. Information professionals spend an average of \$402 per year per user on searching while users spend \$400-\$1500 out of their own pockets.

Healy foresees seven scenarios for the information industry from 2006 on:

1. Direct to consumer publishing
2. Information wants to be free
3. Users as aggregators
4. Open access, open source, open collaboration
5. Libraries morph
6. Social publishing displaces publishers
7. Publishers embrace social publishing

Examples are:

1. [ArtistsFirst](#).
2. [Topix.net](#) (free information subsidized by advertisements).
3. [Stanford on iTunes U](#) (where you can aggregate podcasts).
4. STM leads in open access, open source, open collaboration but publishers are not trying the advertisement-based model for open access, to the surprise of [Outsell](#). An example is the MIT [Registry of Standard Biological Parts](#).

5. Libraries are a thing of the past in the corporate arena. An example of a library that has morphed is [TechXtra](#) at Heriot Watt University.
6. In [Northwest Voice](#), a free community newspaper, the readers are the authors.
7. The [Emerging Technologies Knowledge Base](#) forum supports IT professionals.

Healy closed by drawing an analogy with the Kodak story where the requirements were easy archiving, portable content, ownership of content, easy quality and easy sharing. Eastman's original ideal was to make it as easy to use a camera as it is to use a pencil.

New Players in the Information Life Cycle: Search

Moderator: Judith Russell, Superintendent of Documents, Government Printing Office

Google: Search and the Information Life Cycle

The Google speaker, Anurag Acharya, was unable to leave Paris because of a stolen passport.

National Science Portals; New Partnerships for Global Discovery

[Eleanor Frierson](#), Deputy Director, National Agricultural Library

Frierson began with a quotation from Kevin Kelly in the *New York Times* magazine on May 14, 2006. He described the long term campaign of science to make available the vast, interconnected, peer-reviewed Web of facts. "In science", he said, "there is a natural duty to make what is known searchable". Frierson wanted to show how national science portals might make available "the vast, interconnected, peer-reviewed Web of facts". In 2005 she was invited to speak to a seminar sponsored by the Korean Institute for Science and Technology Evaluation and Policy (KISTEP) because the Korean authorities were considering establishing a national science portal. She considered a number of issues including authority (what are the policies and who sets them?), audience (who are the users and what are their needs?), content (is it on open access, or is it proprietary, or is it both, and is it government information?) and language. She found examples of portals in Australia, Canada, France, Japan, Germany and the United States.

The [Australian portal](#) is federal, delivering science information and services to industry, investors and the research community. The language used is English. The portal features a Web browser. It is funded federally and government agencies participate. The home page is attractive and vibrant. The [Canadian portal](#) is a federal one aimed at "Canadians". It has some fee-based services supplied by the Canada Institute for Scientific and Technical Information ([CISTI](#)). Document delivery is supported by [CISTI](#). The portal is in English and French. The [French portal](#) is federally funded for a general audience. It is in French. Content seems to be free of charge; it is not obvious if there are off-Web services. This is a journalistic-looking site. The [German portal](#), Vascoda, is federal. It is the best funded of the national portals although it was done on a short-term basis. It has German and English versions. It offers federal information plus for-fee services, with Deep Web searching and browsing. Content includes more than 20 virtual libraries and four networks. This is the most ambitious of the portals that Frierson compared. It is similar in appearance to the [Canadian portal](#). The [Japanese portal](#) is run by the Japan Science and Technology Agency, JST. It is mainly in Japanese although there are some English pages. Off-Web services include document delivery and translation. There are some for-fee services. [Postscript: there is now a JST site, [Science Links Japan](#), in English.]

The [US portal](#) science.gov, "FirstGov for Science", is aimed at science-attentive citizens. It is in English. Off-Web services are offered in response to "contact us" e-mails. Sixteen organizations in 12 government agencies participate. Funding is by voluntary contributions. The portal has an association with CENDI, an interagency working group of senior Scientific and Technical Information (STI) Managers from [12 U.S. federal agencies](#): There is also a science.gov alliance. No formal feedback form is offered on the site. Software is under development that will permit the portal to link to up to 1000 databases in the future (no date set).

All these portals have a different “look and feel” and different content. Many are monolingual. Some offer for-fee services: [Vascoda](#), for example, has a mixture of free and for-fee content. [Vascoda](#) is the only one of the portals that seems to want to capture all relevant content including proprietary content. [Science.gov](#) is unique in integrating Web site and Deep Web search functionality.

How could national portals partner? They could:

- acknowledge the existence of the other portals
- facilitate searching across multiple portals
- exchange experience and expertise
- collaborate on facilitating multi-lingual discovery
- establish joint projects, and
- work towards science.world.

[The concept of science.world was presented by [Walter Warnick](#) in a later talk at the conference.]

Some progress has been made. Frierson has contacted [Vascoda](#) staff and discussions will take place at the World Library and Information Congress to be held by the International Federation of Library Associations and Institutions ([IFLA](#)) in Seoul, Republic of Korea, August 20-24, 2006. [Postscript: since this talk was given, [ICSTI](#) has decided to move forward with a project to support better collaboration among national science portals.]

[CrossRef: Interoperability or Market Edge?](#)

[Chuck Koscher](#), Director of Technology, CrossRef

The word “interoperability” sounds upbeat and implies openness and making life easier. “Market edge” may or may not be a good thing. [CrossRef](#) is an independent membership association, founded and directed by publishers. CrossRef’s mission is to serve as the complete citation linking backbone for all scholarly literature online, as a means of lowering barriers to content discovery and access for the researcher. CrossRef’s mandate is to connect users to primary research content, by enabling publishers to do collectively what they cannot do individually. [CrossRef](#) is also the official [DOI](#) registration agency for scholarly and professional publications. It operates a cross-publisher citation linking system that allows a researcher to click on a reference citation on one publisher’s platform and link directly to the cited content on another publisher’s platform, subject to the target publisher’s access control practices.

A Digital Object Identifier ([DOI](#)) is a unique string created to identify a piece of intellectual property in an online environment. A unique alphanumeric string is assigned to a digital object, e.g., an electronic journal article or a book chapter. In the [CrossRef](#) system, each [DOI](#) is associated with a set of basic metadata and a URL pointer to the full text, so that it uniquely identifies the content item and provides a persistent link to its location on the internet.

[CrossRef](#) has moved beyond basic reference linking: cited-by linking is now possible, so authors can ask “Who has cited my article?”. It is also getting active in metadata distribution to Microsoft, EBSCO, Google etc. (CrossRef Web Services). It is a non-profit, member-owned organization with a 16-member board. More than 1600 publishers participate, of which 352 are members. More than 1000 libraries have no-fee use. There are more than 50 affiliates, associates and agents.

A simple text query on the home page of the [Web site](#) allows a user to enter a literature reference and get it parsed into a [DOI](#). This is a free service. Forward-linking (“cited-by”) has been available for about one year. Thirty million cited-by references have been parsed; 43 million citations have not yet been resolved to a [DOI](#). Using [CrossRef](#) gives greater precision than Google: much noise is eliminated. Public Library of Science is using the [CrossRef](#) forward-linking service.

Publishers are paying for [CrossRef](#) and benefiting from it. Smaller companies and libraries (e.g., University of Chicago, University of Birmingham) benefit from using [CrossRef](#). Open URL users also benefit, but the *real* benefit is to the end-user scholars and researchers who carry out 10-12 million resolutions a month. Publishers can get statistics on usage and identify any problems. About 90% of [DOIs](#) in CrossRef relate to journal articles, about 7% to conferences and about 3% to books. There are also [DOIs](#) for tables and smaller items in journal articles. *Using DOIs* is free. [CrossRef's OpenURL resolver](#) resolves references for free.

[CrossRef](#) demonstrably aids interoperability, so what about market edge? A large segment of the [CrossRef](#) community benefits from the service at no cost. Large publishers participate at very high levels and they pay a proportionate cost. Small publishers are enabled to provide a service that they could never have provided on their own. As a [CrossRef](#) board member has said "reference linking has become an essential commodity".

Innovative Relationships: Who's Partnering with Whom?

Moderator: Kurt Mulholm, [ICSTI](#) Honorary Member

Acquisitions, Partnerships and Opportunism in the New Marketplace

[Boe Horton](#), Chief Operating Officer, [CSA](#)

[CSA](#) is a worldwide information company which specializes in publishing and distributing, in print and electronically, over 100 bibliographic and full-text databases and journals in four primary editorial areas: natural sciences, social sciences, arts and humanities, and technology. It is not just an abstracting and indexing company: it also produces [RefWorks](#) an online research management, writing and collaboration tool designed to help researchers gather, manage, store and share all types of information, as well as generate citations and bibliographies. Through recent acquisitions, [CSA](#) now offers over one million researcher profiles through its Community of Scholars product. Researchers can expand their search for relevant information beyond the document record to the relevant set of people and organizations working to increase knowledge in highly specific areas of interest. Community of Scholars provides: authoritative information about more than one million scholars and organizations around the world; relevant and immediate access to researchers identified by research subject; and verified affiliation and publication information in a regularly updated and searchable format. RefWorks and Community of Scholars are part of [CSA](#) but have separate reporting functions.

Horton outlined some reasons for partnership, e.g., to increase market penetration or add strategic value. The partnership must fit the company's mission and provide new value for customers. The customers are not only libraries: the actual user is faculty or a post-doc. Reverse mentoring is being used to look at the "digital native" defined earlier by Leigh Healy. Partnership may aid product development and speed to market: important when disruptive technologies cause such fast changes. Partnerships will also have financial impact.

Partners may be in technology, content, development, distribution, linking or production. They may be government organizations, or non-government organizations, or public or private companies. Examples of [CSA](#) partners are the U.S. Geological Survey National Biological Information Infrastructure (USGS [NBII](#)), [MuseGlobal](#), BioOne and the United Nations. Horton provided some case studies.

[MuseGlobal](#) is a federated search solution. [CSA](#) partnered because there was customer demand. [MuseGlobal](#) could be integrated with CSA Illumina, maximizing the use of customers' key electronic resources. The CSA Illumina interface provides access to more than 100 databases published by [CSA](#) and its publishing partners. It is designed to provide a simple, more user-friendly approach to searching for novice users while maintaining powerful options for users who

require them. There was no reason for [CSA](#) to re-invent the wheel: it licensed the technology from [MuseGlobal](#) and did the integration to produce the federated searching solution MultiSearch.

CSA partnered with BioOne to help expand their presence in markets outside North America and to increase value for users and hence revenue of CSA Illumina in these markets. BioOne is an aggregation of 82 bioscience research journals (and 1 book) from more than 66 publishers. The BioOne bibliographic database is an indexed and searchable collection of abstracts that link to the full text articles available from the [BioOne organization](#). It provides integrated, cost-effective access to a linked information resource of interrelated journals focused on the biological, ecological and environmental sciences. BioOne was developed by the American Institute of Biological Sciences (AIBS), SPARC (the Scholarly Publishing and Academic Resources Coalition), the University of Kansas, the Greater Western Library Alliance, and Allen Press. [CSA](#) is the exclusive distribution partner outside North America. [CSA](#) created an abstracting and indexing front end.

[CSA](#) has worked with the United Nations Food and Agriculture Organization for more than 25 years, creating and distributing Aquatic Sciences and Fisheries Abstracts (ASFA). Input to ASFA is provided by a growing international network of information centers monitoring over 5,000 serial publications, books, reports, conference proceedings, translations and limited distribution literature. The outcomes of the relationship are cost effective creation of vital content to member organizations; efficient global distribution of product in electronic and print format; and free distribution to lesser developed countries.

Horton ended by outlining some of the pitfalls of partnerships, starting with unclear expectations (it is a good idea to consider a three-year framework) and unrealistic timelines. Major channel conflict can be a problem, but a minor conflict is OK. Undefined rules of engagement, e.g., who does U.S. and international marketing, will damage a partnership. The venture may fail if insufficient time and resources are dedicated to it on both sides. Other pitfalls are unclear exit provision, a change in senior management, and changes in staffing (e.g., relationship managers).

The National Biological Information Infrastructure: A Public Private Partnership

[Kate Kase](#), U.S. Geological Survey

The mission of the U.S. Geological Survey National Biological Information Infrastructure (USGS [NBII](#)) is to create a pathway to the data, regardless of who produced them. [NBII](#) was established in 1993 in response to a National Academy of Sciences report. It is a broad *collaborative* program to provide access to data and information on the nation's biological resources, and tools for integration and analysis.

Over \$1 billion a year is invested in biological data collection and management. Coordination and integration are limited and there are few common standards. Access to existing data, information, and analysis tools is limited. [NBII](#)'s objective is to maximize the return on the research and data collection investment for decision making. It is only through careful preservation and continued access to data and information that we can compare the present with the past, sometimes enabling insights into the genesis of what may appear to be newly discovered organisms, organisms discovered in new places, or changes in the status of the environment. There are about 250 partners in the [NBII](#) collaboration, run by the USGS program management office. Data and information from many diverse sources are distributed, enabling people to see the whole picture. The federated approach involves federal agencies, state, local and tribal government organizations, universities, non-government organizations (NGO) and private sector companies. There are many partners and partner types.

Government-to-government partnerships make sure that the taxpayer's dollar is used wisely. An example is the partnership between the US Geological Survey through the [NBII](#), and the U.S. Fish and Wildlife Service. The partnership has a flexible structure and shared goals, leading to

better outcomes. Migratory birds are the first focus. The aim is ensuring that good data and information are available to assist in decision making about the effective management of species.

A government-to-NGO example is [NatureServe](#) and [NBII](#). [NatureServe](#) is a non-profit conservation organization that provides data and information about natural resources, specifically threatened and endangered species. Through this partnership, there are now available through the [NBII](#) more than 1500 species profiles, that is, pictures and detailed descriptions of species of concern. In addition, [NBII](#) is able to provide direct access to many data collected by the Natural Heritage Programs which exist in every state, something that [NBII](#) could not easily do if it had to negotiate with each Heritage program individually. Also, through the NatureServe organization, [NBII](#) has access to NatureServe members as stakeholders and is able to get feedback on user information needs.

Invasive Plant Atlas of New England ([IPANE](#)) is an example of government-to-citizen partnership. Its mission is to create a comprehensive Web-accessible database of invasive and potentially invasive plants in New England that will be continually updated by a network of professionals and trained volunteers. It expands the resources available to slow or arrest the spread of a potentially invasive species.

Government-to-business relationships can promote effective interactions, and expand the expertise and resource pool. An example is the relationship, of many years standing, between [NBII](#) and [CSA](#). It has an innovative business model that allows for cost sharing and tests new market niches and product delivery mechanisms. Examples are the CSA-NBII Biocomplexity Thesaurus and a peer-reviewed, open access e-journal, *Sustainability: Science, Practice, and Policy*.

A second government-to-business example is the relationship of [NBII](#) with [Plumtree Software](#) (now owned by BEA Systems). [Plumtree Software](#) provides enterprise solutions to bridge disparate work groups, IT systems and business processes with its cross-platform portal. The Plumtree portal powers the NBII's intranet portal. This is a collaborative relationship outside a formalized contractual structure. Side-by-side participation provides mutual benefits. Kase ended her presentation with the slogan "building knowledge through partnerships".

Multi-national and Multi-institutional Partnerships

[Gerri Potash](#), Vice President and CTO, [Nerac](#)

[Nerac](#) offers critical research, patents and alerts. It has 250 employees: chemists, biologists and engineers. It started up forty years ago, as New England Research Applications Center, with funding from the National Aeronautics and Space Administration (NASA), and was one of 10 centers throughout the country that addressed the need for private businesses to have access to NASA's research. The relationship with NASA was closed in the late 1980s. The University of Connecticut used to be Nerac's host organization and [Nerac](#) was located on campus until 1987.

During the 1980s and 1990s, Nerac's partnerships were with content providers. In 1985, [Nerac](#) discovered the [British Library](#) through Yellow Pages. [Nerac](#) was looking for "weird and wonderful stuff". A [Nerac](#) client needed full text articles. The relationship evolved into a table of contents service. The [British Library](#) increased its number of document orders through [Nerac](#) tables of contents. This was partnership "out of necessity" for [Nerac](#); more strategic partnerships with [Univention](#), [Elsevier](#), and [Mark Logic](#) followed.

[Nerac](#) found [Univention](#) on the Internet and met the company at Patent Information User Group and London Online meetings. [Nerac](#) wanted to improve its patent research capabilities and was looking for images for Patent Cooperation Treaty (PCT) patents. (PCT patent applications are those administered by the World Intellectual Property Organization.) The evolution of the relationship was key content for [Nerac](#).

Elsevier was somewhat of a chance encounter. The two companies had met in various places. [Nerac](#) was looking for content and Elsevier was looking for revenue. [Nerac](#) wanted to improve its technical research capabilities. The evolution of the relationship was key content, i.e., access to full text. “Nerac.180” was developed in collaboration with Engineering Information (EI), a Reed Elsevier company. “Nerac 180 powered by EI” is 180° opposite to what [Nerac](#) had been offering previously. [Nerac](#) did not have a public search interface until it partnered with EI. Engineers can now call Nerac to get help so that they get precisely what they are looking for.

[Nerac](#), however, was still not so very different from its early technological appearance in terms of “the green screen”. [Mark Logic](#) helped to create the “new [Nerac](#)”. [Nerac](#) is an employee-owned company. Being employee-owned was just a natural progression for Nerac because the staff had, through collaborative work projects such as the Mark Logic design project, come to embrace that taking ownership and personal responsibility is what is required in order to make big changes happen. In a chance encounter at a meeting of the Society of Competitive Intelligence Professionals, a [Nerac](#) analyst ran into a [Mark Logic](#) salesperson. [Mark Logic](#) was looking for clients and [Nerac](#) was looking for a way to integrate content from disparate sources. The strategic reason for a relationship was to allow [Nerac](#) to move from search to research. The evolution of the relationship was to make [Nerac](#) consultative, but “we call them but they call us too” now applies. The relationship with [Mark Logic](#) is probably the most strategic of Nerac’s relationships.

The “engaged” relationships are the ones that made it into the current presentation. [Nerac](#) has had other relationships along the way but the ones described in this talk are the “engaged” ones. There are some local relationships too such as “Nerac Knows” for the Connecticut Technology Council, and the University of Connecticut’s system “Powered by Nerac”. In “reaching out and touching someone”, Nerac also has “not-so-local” involvement, but it has no overseas offices and no sales agents. The analysts themselves pick up the phone to touch someone when it comes to overseas relationships.

Globalization. Internationalization

Moderator: R. Parthasarathy, Co-founder and co-CEO, Lapiz Digital Services

Offshore and New Worldwide Capacities

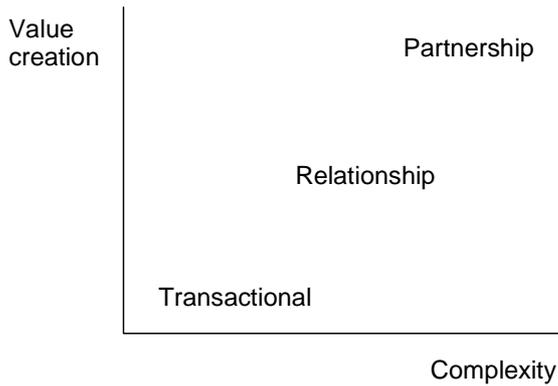
[Prashant Pisharoty](#), [Scope e-Knowledge Center](#), U.S. Operations

Outsourcing is not a new phenomenon. America’s covered wagon covers and ship sales were made in Scotland with cloth imported from India. Much more recently, business processes were outsourced; nowadays financial service providers are outsourcing *core* activities to India. The large pool of trained staff in India now matters more than just cost saving. The global outsourcing market is growing at a Compound Annual Growth Rate of more than 9%. By 2010, the market will be worth \$230 billion. The United States will continue to drive this. There is a move, now, away from business process outsourcing to knowledge process outsourcing.

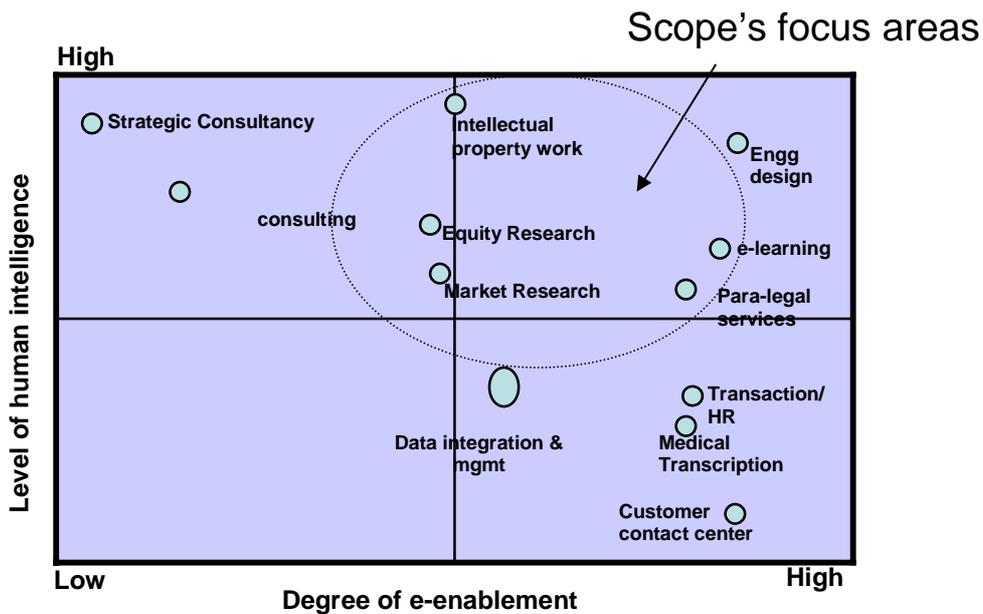
India has many outsourcing advantages, including cost reductions of 40-60%; quality, in terms of highly qualified staff; productivity (the time zone difference gives a company a 24-hour business day) and a high learning curve effect. Multinational companies such as Pfizer, Novartis, AstraZeneca, Microsoft, Intel, IBM, HSBC, American Express and Motorola are off-shoring work to India.

India’s competition includes Ireland, Israel, Russia, Eastern Europe, China and the Philippines. China has the required skills. Language is a problem, but some 200 million people are learning English in preparation for the Beijing Olympics. The Philippines has a cultural affinity with the United States, and a better proportion of suitable employees, but it is a small country with limited resources. India’s competitive advantage lies in having pool of qualified resources at a very competitive price-point.

What does it take to be a player in the outsourcing market? Pisharoty picked three essentials: market, people and knowledge, and process/operations. In the market arena, an Indian (or other) company must (among other things) build a brand, share the knowledge process outsourcing mindset, and earn the client's respect. It must be able to recruit and retain people with analytical and knowledge management skills, and multi-lingual capabilities. It must have the right processes in place to ensure high quality work and delivery, data integrity and security, end-to-end service and continuous improvement. In a typical knowledge process outsourcing model the company must take baby steps up the ladder:



Pisharoty used [Scope](#) (slogan “minds@work”) as a case study. The company has 19 years of expertise with over 1000 assignments from more than 350 clients. It has an excellent pedigree and offers a wide range of services. It is accredited by ISO 9001 and other standards. Its high caliber, young, qualified staff have an average age of 25. It exhibits professionalism and has prior experience. [Scope](#) is a client-focused company with solid information, research and analytical skills. It has three divisions: database services, research services, and patent services. It culls data from a variety of formats and sources, adds value to knowledge, and delivers at consistently world-class standards. [Scope](#) has moved up the value chain from documentation to “abstracting-and-indexing-plus”, to consulting: Pisharoty finished with a diagram showing how [Scope](#) has migrated to higher-end services:



In the question and answer session afterwards, Pisharoty was asked about Scope's clients. They are mostly publishers and intermediaries. He gave an example of re-purposing and re-packaging content, that demonstrated Scope's ability to address different segments, and the company's customization capabilities. The example was of a publishing client, for which [Scope](#) is taking a heavy pharmaceutical tome, and using its encyclopedic information to develop highly-abstracted, bullet-pointed referential diagnosis solutions that are suitable for point-of-care professionals to read over their handheld computers (personal digital assistants, i.e., PDAs).

Global Marketplace for Primary Publishers

[Sally Morris](#), Association of Learned and Professional Society Publishers ([ALPSP](#))

[ALPSP](#) is the association for non-profit publishers. It has 341 members in 36 countries publishing nearly 10,000 journals (about 40% of the world's output). Non-profit publishers are not only society publishers; they are also university presses, associations, research foundations etc. Probably more than 50% of all journals are from non-profit publishers. These publishers have no shareholders and if they make any surplus, it is invested in development and used to support the university, or the society's member programs. They are mission-driven. Some are very old organizations indeed. They are usually specialists publishing a small number of journals with prices about one third of those charged by for-profit publishers. Their articles are highly cited.

Morris spoke particularly about *primary* publishers in this presentation. The first outputs of research are data. Primary articles are the next stage. Secondary publication (bibliographic databases and reviews) follows, although sometimes reviews are seen as tertiary publication. Almost all scholarly journal publishers publish primary articles (although many also publish reviews).

About 2-3 million articles per year are published (ignoring China) in 20,000-24,000 journals worldwide, published by about 6,000 publishers. The market is academic and corporate libraries, but really it is the readers who matter. Some publishers are global organizations (Elsevier lists 73 locations, Oxford University Press has offices in more than 50 countries) but there are thousands of small publishers.

Nevertheless, the *journals* themselves must reach a global market. Some have changed title as a result. For example, the *British Journal of Gastroenterology* is now called *Gut*. An increasing number of manuscript submissions come from outside a journal's home country; online submission systems have helped in this respect. Editors are not necessarily in the home country; often there are editors in more than one country. Editorial Boards are international. The majority of readers may be outside the home country: online delivery means that there is no delay. Researchers are increasingly working internationally in "big science" multinational groups. Although research communities are highly international, they may exclude researchers in less developed countries.

Morris showed a chart showing growth in the number of articles being published. Growth is flat for the USA; it has been overtaken by Europe. East Asia and China are publishing increasing numbers of articles. So, where is the global market not working? Access is now mainly electronic and, in electronic publishing, licensing is the business model rather than a subscription system. The cost per journal is going down. There is high usage of journals to which libraries previously did not subscribe. Separate articles can now be purchased. Figures from the Association of Research Libraries (ARL) show that overall expenditure is still increasing but the cost per journal is going down, as there has been a dramatic increase in the number of serials purchased. According to research by [NESLI2](#) (the United Kingdom's national initiative for the licensing of electronic journals), the average cost to a library per article is less than the cost of ordering the article through interlibrary loan.

In 2004, the Center for Information Behavior and the Evaluation of Research ([CIBER](#)) reported on researchers' views. Researchers claimed to have better access than even before, except in

Central America and Eastern Europe. A study for the [Publishing Research Consortium](#) showed in 2006 that access to the literature was not a significant obstacle to research productivity (with a few exceptions).

Access is being extended. More than 20% of publishers have at least some fully open access journals. Others offer an open access option. Others make content available free after a number of months. Some permit self-archiving, but embargoes are increasingly being introduced. There is some demand (but not much) from the general public for open access, especially in health and environment. More useful, perhaps, is explanation on interpretation of data for non-scientists as in [patientINFORM](#).

Some publishers license local reprinting in less developed countries, to give cheaper access. There are services, such as the Health InterNetwork Access to Research Initiative ([HINARI](#)), that offer free or inexpensive online access but this is only useful if an infrastructure is in place. Unfortunately this sort of initiative creates an impression that articles are free, which is hard on local publishers. Programs to develop local publishers' knowledge and skills are offered by [ALPSP](#), International Network for the Availability of Scientific Publications ([INASP](#)) and others.

Journals are important for features such as selectivity (by peer review), presentation (with the added value of editing etc.), collection (in which the journal becomes an "envelope") and dissemination. Publishers can make a new journal happen and develop and enhance existing journals. Are journals still important in the digital age? Authors think so and libraries think so. However, do readers think so, or are self-archived preprints sufficient?

The growth of self-archiving policies may have unintended consequences. Potential positive consequences are increasing readership and an increasing number of citations. Potential negative consequences are version confusion, lower usage on publishers' sites, leading to library subscription cancellations and knock-on effects for societies' other activities.

The way forward must involve collaboration between funders and publishers, and collaboration between academia and publishers. They should work together to maximize access without undermining what journals do. They should support experimentation with new models. They should respect the need for publishers to remain viable (if readers still *want* journals). Progress can be achieved with time embargoes on free access, links to full text on the publisher Web site, and deposit by publishers of the final version of an article. We can improve science by *working together*.

Experience of Global Responsibilities in Science Publishing: Examples from an Editor-in-Chief for ICSU International Union of Crystallography

[John Helliwell](#), School of Chemistry, University of Manchester, England

The International Union of Crystallography ([IUCr](#)) publishes eight journals including six *Acta Crystallographica* journals. Crystallography Journals Online is accessible globally through the [IUCr](#) and through [Blackwell Synergy](#) and the International Network for the Availability of Scientific Publications ([INASP](#)). The [IUCr](#) Journal Grants Fund pays, or part-pays, for access to Crystallography Journals Online from institutions in Argentina, China, Poland and other countries.

A hybrid open access model was introduced in January 2004. Authors can choose to pay \$900 (£500) to put their article on open access. Free access is provided to certain types of published material, e.g. editorials, letters to the editor, book and software reviews, and obituaries. The Joint Information Systems Committee ([JISC](#)) pays for authors in the United Kingdom to publish an open access article. In the developing world, authors have no-one to pay the open access fee for them. Access is truly global: over 45,000 hosts make more than 1.5 million requests a week for items from Crystallography Journals Online. In a typical week, the journals are accessed from approximately 150 countries.

As an example with interesting global trends, Helliwell described how authors in over 70 countries submit articles to *Acta Crystallographica Section E: Structure Reports Online*, and where there has been a marked increase in submissions from China, Turkey and Malaysia. Indeed China accounts for 40% of all submissions to *Acta Crystallographica Section E*, for 33% of all accepted papers, and for 60% of all rejected or withdrawn papers. In response to the “China challenge”, there are now three Chinese co-editors (out of 45), but more are needed. One problem has been duplication of structure determination because Chinese scientists did not have access to structures already published. Suggestions for notes for authors in Chinese and a Chinese [IUCr](#) office have been considered but discounted. Promotional leaflets have been done in Japanese for the *Journal of Synchrotron Radiation*.

Editors can help authors to improve their work. Helliwell knows, as a practicing scientist, that his own submitted articles are often improved by the peer review process and will thus not be published exactly as submitted. Indeed he would not personally want early versions of his manuscripts to be freely available on the Web as they would simply waste readers' time. In handling around 1000 submitted articles as an Editor over a 15 year period, Helliwell stated that peer review was nearly always a big help to provide an improved, often much improved, publication. Less experienced authors can come from *anywhere* in the world. Helliwell gave a particular example of a paper from a small research institute in a G8 country that had no conclusions section, even after the referees had finished. He suggested to the authors the best emphasis for such a section, a suggestion which they took up, and which they kindly acknowledged. Editorial work like this takes time, and it is worth it, he believes, in terms of the final article quality that is published, but journals are also measured by their speed of publication, including time from submission. As a compromise approach, [IUCr](#) Journals allow approximately two months maximum for revision of a manuscript by authors; if the revisions are substantive or simply tardy then the manuscript can be resubmitted or withdrawn respectively. Publication times have improved markedly since 1999, except for the *Journal of Synchrotron Radiation* which is a new journal and has an emphasis on special issues where one or a few papers can add to publication times for a Journal issue.

Whatever is published is assessed by readers from Nobel Laureates down to PhD students. Data validation is possible before a structure is published: there is a “checkCIF” feature in *Acta Crystallographica Section E: Structure Reports Online* allowing authors to check the accuracy of their Crystallographic Information Files ([CIF](#), the data exchange format used for crystal structures). The end result of all this is a 3-dimensional structure, available online. Helliwell showed an example. Consensus has been reached on the standard for small molecules; there are emerging standards for macromolecular crystallography. *Acta Crystallographica Section F Structural Biology and Crystallization Communications* aims to provide a home for communications on the crystallization and structure determination of biological macromolecules.

Rejection and withdrawal rates across all the [IUCr](#) journals are 20-45%. This is low by the standards of high impact journals and reflects the approach of giving experience and assistance to authors where possible. However, the quality of the scientific argument is paramount and transcends national boundaries, i.e., standards are maintained basically. [IUCr](#) uses its surplus from its publishing operations to fund the Journal new initiatives and funding of projects and travel grants within the IUCr community.

Publishing is not the only field which has been revolutionized by online access. Remote controlled access to instruments is now becoming commonplace. Helliwell believes in universal and equitable access to data and information, in sharing of knowledge, experience and respect across nations, and in full participation in the literature, on an equal footing, from all nations. [IUCr](#) is discussing how to further enhance its efforts in the global Information supply for the good of science. Indeed such principles are now embodied in the new [Global Information Commons Initiative](#). IUCr already makes globally available as much of its data and material as possible within the basic constraint of retaining the sustainability of the IUCr Journals.

DAY TWO. NEW PRODUCTS, IMPACTS AND CHALLENGES

Challenges of the Changing World. Executive Panel

Moderator: [Clifford Lynch](#), Director, Coalition for Networked Information

Panelists: [Sir Roger Elliott](#), ICSU

[Betsy Humphreys](#), Deputy Director, NLM

[Youngsuk \("Y.S."\) Chi](#), Vice Chairman, Elsevier

(In the following,

CL = [Clifford Lynch](#)

RE = [Sir Roger Elliott](#)

BH = [Betsy Humphreys](#)

YS = [Youngsuk \("Y.S."\) Chi](#).)

CL's first question concerned the impact of e-science. Data is becoming an important component of a researcher's needs; it is not only access to articles that matters. How does this change the character of scholarly publishing, and the roles of the various players in the scholarly communications chain?

BH: The human genome project provided a force for change in information use, open access etc. Commercial entities also saw the benefit. New data go in and are immediately connected to old data. Fast recomputation takes place. Many players can link off to their own data. The development of [PubChem](#) is on the same principle and is having the same effect. More than 40 organizations are contributing to [PubChem](#).

[Whole genome association](#) studies have followed the human genome project. These provide large-scale genotype analysis of subjects who have been extensively characterized in clinical studies, such as the [Framingham Heart Study](#) which has for many years been identifying and understanding the factors that increase the risk for cardiovascular disease. Basic associations between genes and patient characteristics are pre-computed so they are not patentable. In large-scale projects of this kind, the National Center for Biotechnology Information ([NCBI](#)) takes the burden of data handling and curation off other organizations. It has a commitment to ensuring persistent access and disseminating the data around the world. Data preservation is less assured when subsets of data from individual experiments are published as supplementary journal articles. Standards are needed around supplemental files. Publishers must add new features for interacting with data. The challenges of maintaining datasets produced by local researchers in institutional repositories offer scope for innovative commercial services.

RE: The physicist's point of view is slightly different. If you mention e-science to physicists they will think of CERN where the Grid was invented. Data must be managed centrally in these Grid operations. Physicists are experts in remote control of instruments. The physical sciences also have large databases where large amounts of data need correlating. Many of these data are generated in small laboratory experiments and it is important that the measurements and analysis be available to other people. We need good analysis tools and standardization. It is here that "information managers", not publishers or librarians in the usual sense, are needed. "Big science will take care of itself. It is the large amounts of small collections of data that need help.

YS: Publishers have no God-given right to support these scientists. From the 1960s to the 1980s there was a vast expansion in the amount of material published. Smaller societies could not cope with this; commercial publishers stepped in. Commercial publishers do not dominate the space nowadays. What Sir Roger said about the small collections is true. The question is: who funds the gathering and curating of the data? The user-pays model is being tested for articles. Will data undergo the same debate? The best discussion takes place before the building of the infrastructures. One size will not fit all. Elsevier originally came up with one big platform [ScienceDirect]. In future Elsevier will need lots of little infrastructures. The world of Web 2.0 is a

different world. Elsevier is struggling with these issues. It is not that Elsevier does not have the necessary technologists.

CL posed a question about open access. This is a world of collaboration and high performance networks. How do we fit the scientific literature into these environments? What happens if the two collaborators do not have the same licenses?

BH: The National Institutes of Health ([NIH](#)) have stated that articles emanating from research that they fund should be made publicly available in PubMed Central but so far this has been a request, not a mandate. A new Appropriations Bill, the text of which appeared today [June 8, 2006] is directing [NIH](#) to make deposition in PubMed Central mandatory. "Free" is very good for the user. It has been said that "free is too expensive for doctors". An attending physician has access to the literature through his or her university library licenses, but nurses, pharmacists, and others on the same health care team often do not. Some 94% of articles in PubMed Central were used in FY 2005. Free is very powerful.

RE: Sally Morris did not tell us what authors want. They want wide dissemination of their work so that they get recognized. Physicists therefore like preprints but they do also want peer review and publication in a peer-reviewed journal. Their colleagues see the preprint and the administration sees the journal article. Too much material is being published at too high a cost. The funding model is at fault. Surpluses from publishing have been used to fund activities of learned societies. The author needs proper repositories. These may be institutional repositories or subject repositories but the latter are preferable. Authors also need search tools. We need a new economic model. Experiments are being carried out but no satisfactory funding model has been found yet. Open access is inevitable: it is what the community needs.

YS: Sir Roger has said it all. No-one wants to restrict access but it is not necessarily true that open access is inevitable. There will be a change in the function of a published article. There is a surge in demand for a new product until the concept becomes boring. Then someone provides a premium for a slightly different service. That in turn causes a new surge in demand. New premium content will appear and will be in a paid-for service. It is not true that the second of two collaborators does not have access: he or she has *paid-for* access. You have to pay for value-added services. All the open access funding policies of governments and other organizations lead to other issues down the road.

BH: The developing world might leap over a technology. There are prototype international mirror sites of PubMed Central: "portable PubMed Centrals". The United Kingdom, Italy, South Africa, Japan and China are trying this. Helping developing countries to avoid large infrastructure costs might be a business for someone.

CL: (to YS) What is the nature of this premium material?

YS: It depends on the discipline. The earlier comments were made in generality. Everything free generates a possibility for a paid-for service. This is how publishing will survive: it has no God-given right to survive.

CL: Two comments. First, time is always the scarcest resource, and that opens up opportunities for premium services that can save time. Second, I was recently reminded that in many fields of scholarship, particularly in the humanities, grants, particularly of the kind we are used to in science, are almost unheard of. We need to remember this in some of the discussions of the economics of scholarly publishing and open access. Final question to the panel: People increasingly want and need to do computations on very large corpora of scientific or scholarly literature. What is this going to do to the scholarly publishing system?

BH: Yes, people want to do this. This is a growth area for Medline licenses. Medline is free, but but there are some restrictions on re-use of information, e.g., if you provide access to Medline

data for others you must routinely update, e.g., to handle retractions. NLM continues to update and maintain the links and connections between original citations and any subsequent retractions, error notices, etc.

RE: Intellectual property rights have been mentioned. If the publisher holds the copyright, there will still be problems with re-use. The European Union has gone ahead with its legislation on *sui generis* database protection. This remains a serious threat to science but to date the legal cases have not involved research information.

YS: I dream of the machine reading more than the human does. Intellectual property is one problem. Whether everything is all in one place also affects the issue. This is very important to the scientist, so we have to find an answer.

CL invited questions from the floor. John Helliwell mentioned microgravity space research and the timing problem, e.g., the tension between release of research data to the scientific community and the desire of the original investigators to be the first to publish based on the data. Data is languishing unused. This is of relevance to “small science” as opposed to marquee projects. BH also talked of the notion of the window in time. There was much discussion of supplemental data standards and the IUPAC International Chemical Identifier ([InChI](#)) was mentioned.

Bonnie Carroll referred back to the consumerism in Healy’s talk. STM information, on the other hand, is sold in its “envelopes”. Will a piece of supplemental information become the unit of doing things in future?

YS: Collaborative filtering (e.g., peer review) is here to stay. The tools used will go through several iterations. When will the funding authorities provide alternative ways of evaluation [for tenure and promotion]? When will there be a move away from impact factors? Frequency of hits and frequency of readings are not enough. Frequency by *quality* is what matters.

Repositories in the Scientific and Technical Information Infrastructure

Moderators: Yuri Arskiy and Igor Markov, VINITI

Global Search and Distributed Repositories

[Walter Warnick](#), Director, Office of Scientific and Technical Information, U.S. Department of Energy (DOE)

The advancement of science depends on the sharing of science knowledge. The Office of Scientific and Technical Information (OSTI) is working to accelerate the advancement of science by accelerating science knowledge diffusion. William R. Brody, president, The Johns Hopkins University, said “Knowledge drives innovation, innovation drives productivity, productivity drives our economic growth.” How do we accelerate knowledge diffusion in multiple disciplines? The “surface Web” is only a small part of the world’s electronic knowledge; in contrast, the deep Web, where the bulk of authoritative science resources is found, is huge. Searching the deep Web presents a challenge. Metasearch, which can be used to search the deep Web, has scaling issues.

OSTI has made progress implementing metasearch through the development of [Science.gov](#), the [E-print Network](#), and [Science Conferences](#). There are 30 databases in [science.gov](#), belonging to 12 separate federal agencies. To make searching them easier, the Office of Scientific and Technical Information (OSTI) is working on metasearch together with the other agencies. [Science.gov](#) was described in an earlier presentation. It searches the U.S. government’s vast stores of scientific and technical information across databases and more than 1,800 science websites and currently accesses over 50 million pages of science information.

The [E-print Network](#) is a gateway to over 20,000 Web sites and databases worldwide, containing over half-a-million articles, including e-prints in basic and applied sciences, primarily in physics but also including subject areas such as chemistry, biology and life sciences, materials science, nuclear sciences and engineering, energy research, computer and information technologies, and other disciplines of interest to DOE. It has a set of specialized tools and features designed to facilitate the exchange and use of scientific information, including a unique Deep Web search capability that combines full-text searching through PDF documents residing on e-print Web sites with a distributed search across e-print databases.

[Science Conferences](#) is OSTI's Science Conferences portal. This distributed portal provides access to science and technology conference proceedings and conference papers from a number of authoritative sites (mainly professional societies and national laboratories) whose areas of interest in the physical sciences and technology intersect those of the Department of Energy.

Now OSTI, working in cooperation with international partners, has developed a prototype that which could be a model for an eventual "Science.world," thereby taking this technology to a global scale. OSTI is the operating agent for the Energy Technology Data Exchange ([ETDE](#)), an international energy information exchange agreement formed in 1987 under the International Energy Agency (IEA). Its mission is to provide governments, industry and the research community in the member countries with access to the widest range of information on energy research, science and technology and to increase dissemination of this information to developing countries. There are 16 member countries.

ETDE World Energy Base ([ETDEWEB](#)) is an Internet tool for disseminating the energy research and technology information that is collected and exchanged. It includes information on the environmental impact of energy production and use, including climate change; energy R&D; energy policy; nuclear, coal, hydrocarbon and renewable energy technologies and much more. It also includes a distributed searching option for one-stop searching of related science sites. ETDE's information can also be found in commercial host products and in some country-specific databases. ETDE has built the prototype for taking distributed searching to a global scale.

The prototype for Science.world has a metasearch tool for searching numerous international sources. Hits can be sorted by source and by relevance. Challenges are scalability, different languages, ongoing viability of Deep Web resources, and including "user authentication" systems. The U.S. Department of Energy is a willing collaborator in pursuing the vision of Science.world, and welcomes other interested parties to begin a dialog and a plan of action for making this vision a reality. [Warnick](#) offered demonstrations of the Science.world prototype to anyone wanting to share in the vision.

Towards a French Repository: Hyper Article on Line (HAL)

[Herbert Grutemeier](#), Institut de l'Information Scientifique et Technique ([INIST](#))

[CNRS](#) is the biggest research organization in Europe. The Institut de l'Information Scientifique et Technique ([INIST](#)) is an institute within [CNRS](#). [CNRS](#) has a strategic role on the French road to open access. La Direction de l'Information Scientifique (DIS) was created in June 2005 and the Director General of [CNRS](#) launched a steering committee for open archives ("comité de pilotage des archives ouvertes", CPAO) in March 2006.

Besides [CNRS](#) there are several smaller national bodies such as the medical research institute, Institut National de la Santé et de la Recherche Médicale, INSERM; the computer science research institute, Institut National de Recherche en Informatique et en Automatique, INRIA; and the National Institute for Agricultural Research, INRA. France also has about 90 universities and more than 200 Grandes Écoles. Many laboratories are joint ventures. This is still a very centralized system. It could be that for this reason the French situation is exceptional with regard to open access policies and achievements.

[CNRS](#) and INSERM were among the initial signatories of the [Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities](#), along with 17 other international research and cultural heritage organizations. [CNRS](#) partners with the open access publisher BioMed Central. In 2000, [CNRS](#) created Centre pour la Communication Scientifique Directe, [CCSD](#) for direct communication between researchers, and in 2001 CCSD developed the Hyper Article on Line ([HAL](#)) server. HAL's software provides an interface for authors to upload into the CCSD database their manuscripts of scholarly articles in all fields.

Following the successful [conference in Paris](#) in January 2003, organized by [INIST](#), ICSTI and others, [INIST](#) maintains an [open access Web site](#). CNRS researchers were also involved, on behalf of the Civil Society, in the World Summit on the Information Society ([WSIS](#)), enabling open access issues to appear in the WSIS official documents.

[INIST](#), INSERM, INRIA, and INRA are involved in the archive project overseen by CPAO. It is driven by three factors: the [arXiv](#) concept of direct scientific communication; the move to open access as in the "BBB" declarations (Budapest, February 2002, Bethesda, April 2003, and Berlin, October 2003); and institutions' needs for a detailed panorama of publications. Gruttemeier listed some agreed principles. The [green road to open access](#) (self-archiving) is the shortest. There is a need for visibility and monitoring of scientific output. There is also a need for gaining independence from commercially-driven publishing. Publication costs should be part of research costs. Maximum impact is achieved by maximum access. There is an advantage to one, unique system.

Gruttemeier outlined the CNRS general policy for an institutional repository. Institutional repositories give leverage to open access. HAL is a unique archive with one, single infrastructure, but several views on the data. The policy has two levels: making contribution of at least metadata to the archive mandatory (although the word "mandatory" is now out of favor) and encouraging full-text open access contributions. Institutional repositories are increasingly important in the research assessment process. There is an emphasis on quality and services. This goes beyond the Open Archives Initiative ([OAI](#)) as far as data description and representation are concerned: more standardization efforts are needed.

Scientific communications involve both research papers and research data. The [Berlin Declaration](#) covers both but this presentation covers only papers. [HAL](#) contains about 28,000 full-text documents. There is a strong link between [HAL](#) and [arXiv](#) but HAL has richer metadata than [arXiv](#). It is a full text, multidisciplinary archive for all kinds of scientific publications. It employs the principles of self-archiving (with editorial assistance by librarians and information professionals) and long term archiving. Standards used include [ReDIF](#), a metadata format used to describe academic documents, their authors and institutions, the Open Archives Initiative Protocol for Metadata Harvesting ([OAI-PMH](#)), and [RSS](#) feeds. Documents in disciplines covered by [arXiv](#) are automatically transferred from HAL to [arXiv](#).

The role of the CPAO is to apply rules of good practice, support development tools and inform and communicate on open archives among all disciplines, communities and networks. [HAL](#) is both institutional and international. An agreement at a national level has been reached among the four initial partner institutions CNRS, INSERM, INRIA and INRA, joined by Cemagref, [CIRAD](#) (a French institute for agronomic research in development of the southern nations), Institut Pasteur and the Institut de Recherche pour le Développement ([IRD](#)) as well as (and this is of high political importance) the Conférence des Présidents d'Université (CPU) and the Conférence des Grandes Écoles (CGE). One of the cooperation issues among these partners was the elaboration of a "guide book" for researchers, focusing on legal aspects and providing information about initiatives like the [SHERPA/ROMEO](#) directory of publisher policies. ([SHERPA](#) stands for "Securing a Hybrid Environment for Research Preservation and Access").

[HAL](#) allows for versioning. Different views onto [HAL](#) can be created, for example, a university or research unit may be shown its own articles. Physics and mathematics are the main categories at present. In future a terminology portal developed by INIST will be linked to [HAL](#).

[DSpace](#): Institutional Repositories

[Ann Devenish](#), Library of the Marine Biological Laboratory and Woods Hole Oceanographic Institution

Woods Hole Open Access Server ([WHOAS](#)) is a [DSpace](#) institutional repository. There are several Woods Hole science institutions scattered across the town of Woods Hole, Massachusetts (on Cape Cod): MBL, formerly the Marine Biology laboratory; the National Marine Fisheries Service/Northeast Fisheries Science Center; the Sea Education Association (SEA); the U. S. Geological Survey (USGS); the Woods Hole Oceanographic Institution (WHOI); and the Woods Hole Research Center (WHRC). The one MBLWHOI library serves all these institutions from three physical service points. There are also virtual service points, on every computer in the IP range, offering over 98 databases and more than 1300 e-journals.

“Quiet” dreams of an institutional repository began in 2002. Conversations were held with users about the value of an institutional repository and the value of open access. After that, the dreams were shaped by reality. The institutional repository had to be built within existing budget dollars and staff. The idea was to use a “herbarium” model and open source software. Participation by authors was to be voluntary: there were no institutional directives mandating self-archiving. The archive had to support all types of digital formats. The content had to be “Googleable” and openly accessible. The initial focus was on MBL and WHOI authors and content.

A pilot project to build the institutional repository ran from December 2003 to July 2004. The project team needed to decide on the platform (i.e., [eprints](#) or [DSpace](#)), the metadata scheme, workflow and DOI deposit. [DSpace](#) was chosen. It has a number of advantages from a librarian’s point of view, one of which is that it is free. It is an open source platform focused on long term preservation using the [CNRI Handle](#) system. It supports qualified Dublin Core metadata and the Open Archives Initiative protocol for metadata harvest ([OAI-PMH](#)). It has full administrative capabilities; decisions can be made at institutional and departmental levels. From the user’s point of view, it has a Web user interface, it is designed to handle all kinds of digital formats, and metadata input requirements are minimal.

The principle for growing [WHOAS](#) was “keep it simple” (“KISS”). From the IT point of view, the platform has minimum customization, no software development has been carried out, and migration to new versions is only to be done when necessary. For researchers, KISS means that content is “as is” and mediated input is done by library staff. The library staff have no control over author names (they go in as they appear in the work) or over subject terms (so an abstract has lots and lots of words). The total number of metadata records was 40 in July 2005 and 84 in November 2005, but then rose steadily to 928 in May 2006.

Of the 929 metadata records on June 1, 2006, 286 are from content from the International Association of Aquatic and Marine Science Libraries and Information Centers ([IAMSLIC](#)) and the rest are from Woods Hole, namely from 276 technical reports, working papers, and theses, 200 published articles, 140 preprints, 19 books, 6 presentations (or “other” content), and 2 data sets. MBL and WHOI have published 1250 articles since January 2004 in approximately 380 journals. This represents 85% of known Wood Hole science publications, many in journals that permit self-archiving in institutional repositories of preprints or post-prints.

Both “push” and “pull” techniques have been used to capture more content. Push (i.e., voluntary contributions) has produced less than 3% of the total Woods Hole content in [WHOAS](#). For comparison, note that a [report on the NIH Public Access Policy](#) in January 2006 estimated that less than 4% of the total number of relevant articles had been added to PubMed Central. So “pulling” seemed to be the best way ahead. Harvesting (from the sites of publishers who allow

self-archiving) has produced 59% of the Woods Hole articles in [WHOAS](#). Library staff identify, verify, solicit and load other content. The response rate to this solicitation exercise was 27% in 2005, rising to 40% in 2006. Some of Woods Hole's own content is also being digitized: more than 3500 reports published since 1941 and about 1150 theses, from WHOI or WHOI/MIT, published since 1968. The initial focus of the digitization program is on material published since 1990. Some other "on demand" content has also been acquired.

[DOI](#) deposit is being implemented to help expose the content to its widest possible audience. Since [DOI](#) is being broadly accepted in the scientific technical and medical publishing (STM) community, [DOI](#) deposit facilitates linkage to [WHOAS](#) objects by STM publishers. [DOIs](#) are deposited with [CrossRef](#) for original works but not for preprints or published versions of articles, in order to avoid duplication.

Devenish listed some lessons learned: involve your library and IT colleagues early and often; be flexible, patient and persistent; expect the unexpected; be prepared to undo and redo, know your limitations, accept that mistakes are okay; and look forward. In future, she expects "more of the same", e.g., conversations regarding the value of an institutional repository, and the value of open access. There will be more outreach, to recruit content from SEA, USGS, and WHRC.

Moving up the Value Chain: Information for Decision Making

Moderator: [Wendy Warr](#), Owner, [Wendy Warr & Associates](#), IUPAC Liaison to ICSTI

Visualization as a Tool for Analysis

A speaker from [Grokker](#) did not attend.

The Changing Business of Patents: [EPO](#) [Gérard Giroud](#), European Patent Office ([EPO](#))

A patent for an invention is granted to the inventor, giving him or her the right for a limited period to stop others from making, using or selling the invention without the permission of the inventor. When a patent is granted, the invention becomes the property of the inventor, which, like any other form of property or business asset, can be bought, sold, rented or hired. New ideas that are patented lead to the creation of yet more new ideas. Patents thus promote the advancement of science and this is one reason for giving the inventor a monopoly.

Patents are territorial rights; they are granted by national routes, regional routes or the Patent Cooperation Treaty (PCT). No world patent or European Union (Community) patent exists. The number of patents continues to increase (and the number of bad patents increases also), but it is actually the number of countries chosen by an inventor that causes proliferation of patents. Many documents describe the same invention because a separate patent is required for each country in which the inventor wants a monopoly. It would be good if just one examination could be done for all countries but that is not currently possible. Patent offices are major publishers. When you are doing preliminary research for a new project you *must* look at the patent literature. Bridges between the various repositories of scientific publications and patent databases would be desirable.

A monopoly is granted for 20 years but the inventor must pay a fee every year. This encourages people to patent only those inventions that are valuable and useful. Patent information is free to read. Giroud showed a diagram of the flow of applications between trilateral blocs (i.e., the United States, Japan and the 31 European Countries of the EPO). The Japanese are patenting abroad more and more. U.S. patenting in Japan is decreasing.

Patent offices have to accumulate huge collections of scientific and technical documents in order to do the examination of patents. Publishers and universities should note: (a) patent offices are interested in reading their publications and are therefore potential subscribers (b) the patent offices have enormous documentation resources, beyond the patent documents, which may be of great interest for scientists and (c) much cooperation with the publishers is possible.

There is now an urgent need to access Chinese publications since there has been a 35% increase in Chinese patents filings in the last year. This is a problem for [EPO](#). Forty percent of EPO's documents are in Japanese. In future, Chinese scientific literature and patent documents will matter a great deal. Overlooking Chinese prior art could result in an invalid patent, not only in China, but also in other parts of the world. Companies now need to patent in China. In 2004 19.2% of patents filed in China were foreign. Japan is the most active country patenting in China. Applications in Korea and Taiwan and the corresponding published patent documents are also a problem for [EPO](#).

The EPO's headquarters are in Munich. It also has a branch in The Hague and sub-offices in Berlin and Vienna, as well as a liaison office with the European Union in Brussels. The European Patent Organization currently has [31 member states](#). These states have a total of 500 million inhabitants, greater than the population of Japan, but less than that of China. The first function of [EPO](#) is to grant patents. Its second function is to distribute patent information. [EPO](#) had 100 staff in 1978; it has 5918 today. China is now a major patent office. Nearly fifty percent of EPO's patent applications are from Europe, 26% from the United States, 17% from Japan, and 7.5% from the rest of the world.

Giroud listed three challenges for the patent system in the Europe: the cost of European patents, the lack of any unified litigation system, and the fact that there is no Community patent for the European Union. It costs about €31,000 to patent an invention in eight European countries, €10,000 in the United States and €16,000 in Japan. Europe covers 31 countries and multiple languages. The breakdown of the cost in Europe is national maintenance fees 29%, EPO fees 13%, translation 38%, and professional representation 20%. There are three EPO official languages: English, French and German. The application must be submitted in one of these languages and claims must be translated into all languages once the patent is granted. The "London Agreement", signed in 2001, aimed to reduce costs by 45% but, unfortunately, not all countries have ratified it. France, for example, has not.

There has also been agreement over the litigation problem but it is still not ratified. There have been attempts to establish a Community patent. This has been discussed for 30 years but there is no agreement on languages, or on a unique litigation organization. Discussions are also ongoing on scope of protection: whether biotechnology, software and business methods can be patented.

Patents are published 18 months after they are first filed; patents have always tended to be published before any related article in the scholarly literature. [EPO](#) has accumulated an *enormous* number of documents: 56 million patent records in 90 databases, including 54 million records with a searchable abstract. There are 16 million full text records and 55.6 million facsimile documents. These can be accessed through [esp@cenet](#). The system is free of charge and can be interrogated in all applicable languages.

Information Creators in the Face of Advanced Search: the [DTIC](#) Experience

[Paul Ryan](#), Administrator, Defense Technical Information Center ([DTIC](#))

[DTIC](#) has a mission to provide a single point of access for defense scientific and technical information. It centralizes collection and secondary dissemination of scientific and technical information; controls access for information products and networks; balances open access against Department of Defense (DoD) needs; oversees the management of 11 DoD Information Analysis Centers (IACs) that are mandated to provide technical and analytical services to areas

of critical DoD interest; and provides Web services to over 100 DoD Web sites, including the official DoD site, [DefenseLink](#), and various DoD sites related to research and development.

IAC customers include researchers, engineers, scientists, and program managers from DoD, government, industry, and academia. Individuals and organizations registered for [DTIC](#) services are 26% industrial, 25% army, 14% navy; 16% air force, 10% DoD and 7% educational organizations. [DTIC](#) serves as the central collection and dissemination point for DoD's scientific reports: over 2 million collected since DoD's inception in World War II, with an annual acquisition of over 25,000-30,000 new technical reports and summaries of ongoing research. There are more than 170,000 descriptions of independent R&D projects. Ongoing and planned R&D project information is garnered from both the DoD community and industry: defense agencies, DoD laboratories, research and engineering (R&E) contractors, universities and military schools, NATO, etc. Fifty-one percent of the information is unclassified, i.e., access is unlimited; 41% is unclassified but access is limited; and 8% is classified. The content is safeguarded by an authentication system.

[DTIC](#) supplies value-added information expertise: metadata and taxonomies are applied. [DTIC](#) preserves DoD's information assets. [DTIC](#) staff and IAC analysts have advanced degrees and knowledge of DoD information assets, sources and history. Everyone can search DTIC's publicly accessible collections and display or download scientific and technical information, using the [Public STINET](#) service. Web pages supported by [DTIC](#) include [DefenseLink](#), DefendAmerica, Federal Voting Assistance Program, Iraqi Virtual Science Library, and Joint Chief of Staff.

Ryan showed the home page of DTIC's R&E portal. Too many people are going to the Internet for their information: "Googlization" is rife. This is not the way to work if you are doing research for DoD. Some people are paying for information that they could get free from [DTIC](#). Ryan demonstrated the power of searching in the [DTIC](#) environment. Much DoD information is missed by standard Internet search engines. The [DTIC](#) search interface has three versions: [Public STINET](#), STINET for information professionals and a classified level. [STINET MultiSearch](#) allows users to search multiple databases with which they are not familiar. There are a quarter of a million hits a month on this page.

The [DTIC](#) Open Archive Initiative was launched in 2005; 11,000 items were accumulated plus 25,000 more in May 2006. [DTIC](#) is working with Google and Yahoo to bring [DTIC](#) hits to the top of the list (when relevant). Ryan described the results of a comparison between [DTIC](#) search and Google. Various topics relevant to DoD were searched. [DTIC](#) resources include private STINET etc. Google has Google, Google Scholar and Google Print. Results were analyzed by information type, database or library, language, and availability.

There were more hits in the Google search but, in the top 20, the [DTIC](#) hits were the more appropriate. Ryan admitted that this was not a rigorous study and it was not independent, i.e., it was done by [DTIC](#). He tabulated some 2005 results:

	DTIC	Google
Number of highly relevant hits	197	24
Number of hits of medium relevance	230	92
Number of hits of low relevance	195	465
Number of irrelevant hits	113	177
Number non-publicly available	369	0

Ryan also showed a breakdown of hits by source and type. The comparison was repeated in 2006, since the [DTIC](#) content has by now been exposed to Google. The results were mixed (perhaps the comparison was carried out too soon) but there were no additional [DTIC](#) technical reports identified in the Google databases.

Outsell has shown that engineers are now relying less on the open Internet. OCLC has harvested 5874 records from [DTIC](#) for inclusion in WorldCat. Links and handles are being added. [DTIC](#) redesigned the DoD Web site. This work was done by information professionals and not by IT professionals.

National Centers: New Roles, Relationships and Opportunities for Partnership

Moderator: [Ellen Herbst](#), National Technical Information Center ([NTIS](#))

In an earlier session, Bonnie Carroll had speculated about how much of the new technology, and new approaches to information dissemination, would stick. With reference to Bonnie's question, Herbst thought that the new technology, expectations and approaches were indicative of a permanent sea change that we have to accommodate. The economist Josef Schumpeter coined the phrase "creative destruction". In "creative destruction", some things are destroyed to make way for new things. We are experiencing a changing rate of change. User expectations are changing. Using expressions from *The World is Flat*, Herbst said that globalization is at stage 3.0 where 1.0 is globalization of countries, 2.0 is globalization of companies and 3.0 is globalization of individuals. Individuals can now participate globally.

Each of our organizations is part of an ecosystem. All organizations have a place in the balanced ecosystem. We need to re-evaluate our own roles in the system. The four people in this session represent four national centers. They need to partner with others in the ecosystem, looking for partners who complement their competencies. [NTIS](#) runs on a fee-for-service model.

Canada Institute for Scientific and Technical Information ([CISTI](#)) [Bernard Dumouchel](#), [CISTI](#)

Scientific and technical information (STI) centers do not meet so ICSTI was a useful vehicle for information exchange among the centers. The roles of these centers are changing. What is the value proposition? Dumouchel outlined the 2006-2011 Strategy for the National Research Council of Canada ([NRC](#)). This is a research-oriented strategy, not a policy one, as it might have been in the United States. NRC's vision is to be valued as the world's best national research and innovation organization. Its purpose is to provide integrated science and technology solutions in areas of critical importance to Canada. Key messages are partnership and leadership. [CISTI](#) has a national mandate as Canada's library but is also the departmental library of the [NRC](#). It has a large virtual library. It is also a publisher.

Dumouchel showed a chart of the strategic framework. Exploitation matters: information that is *used* adds value. Key factors are vision, mission and outcome; outcome is a new factor. [CISTI](#) wants scientific, technical and medical (STM) information to be recognized for its significance to Canada.

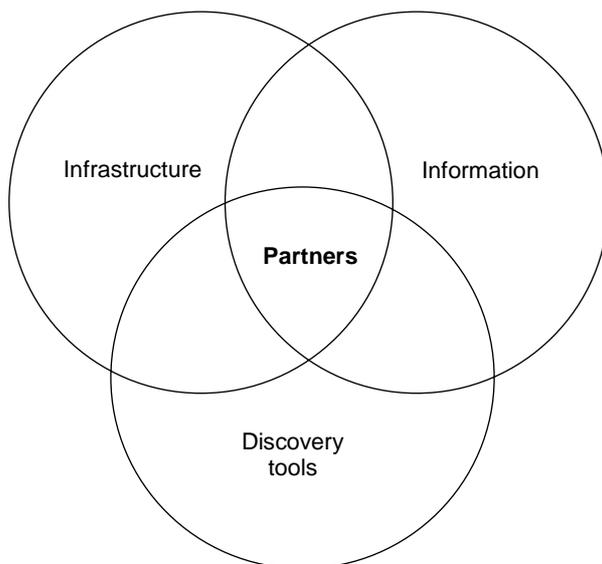
Partnership involves shared or compatible objectives, shared authority and responsibility, joint investment of time and resources, shared risks and benefits, and an explicit agreement that sets out the terms of the partnership. *Partnering* is more like collaboration. Benefits of a partnership are the leverage given to scarce resources; adding value to the partners' respective services; reducing costs; achieving together what one partner alone cannot achieve; and gaining expertise and knowledge.

Dumouchel listed some [CISTI](#) partnerships at work. In 1976, [CISTI](#) and [NLM](#) formed a partnership that continues to this day. In the 1990s, the National Library of Canada and [CISTI](#) offered ROMULUS, a union catalog of STM titles on CD-ROM. That product is now obsolete. In this century, [CISTI](#) is co-publishing monographs with university and society publishers. Most recently, a secure desktop delivery agreement has been negotiated with STM publishers.

The value proposition is as follows. [CISTI](#) provides extended access to the world's STM information through partners, not from its own collection first, as in the past. [CISTI](#) is Canada's foremost publisher. [CISTI](#) has an integrated "infostructure" and supplies services to support commercialization (i.e., competitive intelligence, analysis services etc.). Document delivery partners are the British Library Document Supply Centre ([BLDSC](#)), the Canadian Agriculture Library, [INIST](#), the Institute of Scientific and Technical Information of China ([ISTIC](#)), [KISTI](#), Science and Technology Policy Research and Information Center ([STPI](#)) in Taiwan, [SUNMEDIA](#) in Japan, [NLM](#), and the [University of Alberta](#).

Publishing partners are the Canadian research community, the Canadian government ([CISTI](#) has a repository program for open access in Canada), international agencies such as [HINARI](#), and the [Charlesworth Group](#) (for market reach to China). Online Submission and Peer Review (OSPRey) is a system resulting from an ongoing partnership between [CISTI](#) Publishing and the publishing arm of the Australian Commonwealth Scientific and Industrial Research Organization ([CSIRO](#)).

Canada's scientific "infostructure" (Csi) is a combination of information and infrastructure. The first need of all is a secure, reliable technological infrastructure. The information is the digital repository: [CISTI](#) needs more content for this. [CISTI](#) is also looking for discovery tool partners.



The power of partnering will be the strength behind the infostructure. Csi is a strategic alliance of federal science and technology libraries, the University of Montréal, and Library and Archives Canada. [CISTI](#) also has competitive intelligence partnerships. The [NRC](#) has an industrial research assistance program (IRAP) for small and medium enterprises. Sirsi Dynix has developed a competitive intelligence tool jointly with [CISTI](#).

Dumouchel concluded by listing the outcomes of partnering. OSPRey has accelerated the publishing process, maintained competitiveness, and increased productivity in the editorial office. Csi will provide equitable access for Canadians regardless of sector or geography, assure the preservation of authoritative, high quality and trusted STM information, and help in the acceleration of research and innovation. The competitive intelligence service provides customized, focused services to SMEs, ensures increased competitiveness of Canadian SMEs, and facilitates increased commercialization of R&D discoveries. [CISTI](#) is serious about partnerships and has created a partnership development office.

Korean Institute for Scientific and Technical Information ([KISTI](#))

[Hee-Yoon Choi](#), Director, Knowledge Asset Team, [KISTI](#)

As an institute, [KISTI](#) ensures that Korean researchers and entrepreneurs in the academic, industrial, and public sectors have access to the STM information needed to support their research and development activities. This function is just as the same as the other STM service institutes in the world. However, [KISTI](#) does more than just provide resources: it provides various services from basic information to supercomputing.

The national scientific and technical information activities are driven and operated by public agencies such as [KISTI](#), Korea Education & Research Information Service (KERIS), Korea Institute of Patent Right Information (KIPRIS) and others. Korea is a consumer of information rather than a producer. Most of scientific and technical information resources are written in English. Some organizations, such as Korean Studies Information Co., Ltd., and foreign information agencies, take part in this business, but their capacities are limited. The participation of the private sector is limited due to the domestic market size the characteristics of the information. The government has announced a new policy called the "Promotion Plan for R&D Servicing Businesses" which aims to promote national and private R&D servicing businesses through the private sector.

During the 1990's, information policies supported diversified systems, constructed, operated and serviced independently through role-sharing from organizations of different government agencies. Since the beginning of the 21st century, those policies have turned to an integrated system for national information, which is connected and shared by a network of diversified national information resources. [KISTI](#) provides scientific and technical information, and supercomputing and networking services to users, while KERIS focuses on education and research information including cultural science and social studies, and KIPRIS provides information on industrial property rights. Eighteen centers for specialty research information handle R&D information in specialty fields such as construction, medical science, chemistry etc.

Choi presented a brief history of [KISTI](#). Korea Scientific and Technical Information Center (KORSTIC) was founded in 1962. It was re-established as Korea Institute of Industry and Technology Information (KINITI) by the Ministry of Commerce, Industry and Energy in 1991. Another organization, Korea Research and Development Information Center (KORDIC) was founded by the Ministry of Science and Technology in 1993. KINITI and KORDIC were consolidated by the Korea Research Council of Public Science and Technology in 2000 and eventually formed into [KISTI](#) in 2001. [KISTI](#) was designed to become Korea's top national science and technology information organization, combining all the information-related duties that were previously scattered among different agencies.

The mission of [KISTI](#) is to promote R&D, serve the scientific and technical information infrastructure and contribute to national science and technology and public welfare promotion. [KISTI](#) has three major functions (a) to develop an information portal (b) to analyze value-added information, and (c) to establish a supercomputing infrastructure. As a national scientific and technical information center, it develops, collects and manages information, builds and operates the National Technology Information System (NTIS), supports R&D and distributes high-end information technology. In its role of R&D information analysis support, it carries out technology and industry analysis, and supports value-added information on national strategic technology. As the national supercomputing center, it manages, operates and services supercomputing, builds and operates the national research network, supports R&D and distributes application technologies.

[KISTI](#) has three divisions and four centers that serve the three above functions. It also has three other administration divisions. Offices are in Seoul and Daejeon. [KISTI](#) has 331 full time researchers, most of whom have at least a master's degree. The budget is \$93.8 million. The level of government support has increased from 54.4% in 2004, to over 61.6% in 2006.

The basic strategy is to provide a core infrastructure for supporting a successful national innovation system including the national portal for scientific and technical information, a national R&BD analysis system, the National Technology Information System, the national e-science environment, a global R&D network, and the national scientific and technical information security system that was introduced in 2005. Choi addressed each of these six tasks in turn.

The national portal for scientific and technical information is a knowledge portal giving access to more than 20,000 journals and 50 million documents, including proceedings, research reports, patents etc. through the integrated search site for scientific and technical information. In the national R&BD analysis system, [KISTI](#) is expanding its analysis in technology, trends and the market place into national strategic areas. It offers life-cycle support on national R&D projects with in-depth analysis reports. In the National Technology Information System, [KISTI](#) is looking for new technology trends and ideas that can be expanded from simply planning to the actual commercialization of new technology. This is an open system connecting 20 ministries and 60 institutes.

The national e-science environment allows researchers to share information resources and hardware devices in a cyberspace based on a high-speed network, using a 50 TFlops supercomputer (one of the world's top seven supercomputers) and application technologies. [KISTI](#) is building a cutting-edge, global R&D network for R&D cooperation in the Global Ring Network for Advanced Application Development (GLORIAD) project, in conjunction with the United States, Canada, China, Russia, and The Netherlands. The sixth task is to monitor information and system security of major domestic research institutes operating in a science and technology security center.

Choi next described some major resources in the national portal for scientific and technical information. To establish the national science e-library, [KISTI](#) has constructed a database of over 27 million items from overseas including academic journals, proceedings, and research reports. Academic Society Home Village contains domestic research information and supports the computerization of academic societies. [KISTI](#) collects the publications from societies and constructs databases to manage them systematically for the perpetual conservation and use of domestic research information. The Article Contribution Management System (KISTI-ACOMS) allows researchers to upload their papers on a common site for immediate peer review and quality control. This system is currently being used by 100 academic societies. "yesKISTI" is an integrated search site for science and technology information, providing databases of journal articles, proceedings, research reports, patents, analysis, trends, manpower and seminar videos. yesKISTI allows researchers to search all these different databases with just one interface.

[KISTI](#) offers a document delivery service, supplying 300,000 documents a year, satisfying 95% of user requests. Out of the total number of documents delivered in 2005, 84% were from [KISTI](#) holdings. The national digital science library (NDSL) is another information knowledge portal for science and technology information. It is a national scholarly portal for publications and a nationwide automatic resource sharing system. It focuses on journals and proceeding papers. Over 40% of the resources provided by NDSL concern applied science and 16.5% natural sciences. NDSL is a one-stop, comprehensive search and full-text supply service for 32 million international papers from 49,270 scholarly journals and 178,465 proceedings. It also provides a one-click gateway service to 13 million electronic full text articles in 18,000 core e-journals.

The Korean Electronic Site License Initiative (KESLI) is a library consortium for the joint purchasing of e-publications. A total of 404 libraries (in private companies, universities, research institutes, medical institutes, and public bodies) participate, including 318 in the e-journal consortium. In the e-journal consortium, 14,716 e-journals are licensed under contracts with 73 publishers. Two hundred and fifty-nine libraries participate in the inter-library loan consortium. There is no legal or institutional apparatus supporting the organization and operations of this consortium, but KESLI has been growing and developing on demand. Member libraries of the

consortium buy publications and share them, through the NDSL site, with general users on a pay-per-view basis.

The national R&BD analysis system promotes the efficiency of R&D activities by providing information analysis on domestic and foreign science and technology trends, predicting future research arenas and technology markets. Over 17,000 reports are written annually, on technology commercialization, nanotechnology, trends, future strategies, technology, and industry information. There are four parts to the national R&BD analysis system: human networks which bring together a pool of the finest brains in the fields; advanced systems which systematically discover new promising trends; R&BD analysis which integrates competitive technology intelligence analysis and business feasibility analysis; and an infrastructure for analysis using databases of journals and patents with supercomputing power. One of the human networks is Retired Scientists and Engineers Advancing Technology (ReSEAT), a brain-pool of retired scientists and engineers. [KISTI](#) uses "silver power", R&D know-how, and the accumulated knowledge of retired scientists and engineers for trend alerting, information analysis and R&D consultation.

As a metaphor for the analysis process at [KISTI](#), Choi used the movie, *Minority Report*, starring Tom Cruise. *Minority Report* is about a cop of the future, working in a division of the police department that arrests killers before they commit crimes, thanks to future-viewing technology. Three elements are needed in order to catch criminals before they commit crimes: precogs who foretell the crime, future viewing technology, and pre-crime cops who analyze the crime scene. In [KISTI](#), the precogs are human networks who give in-depth analysis and real-time consulting, and analysis systems with high performance computers, and bibliography, patent, and bio-information databases. The future viewing technology is *Minority Report*. The pre-crime cops are the experts at [KISTI](#) who analyze based on-demand, develop tools and manage the analysis systems. [KISTI](#) has a pool of minds working together with its own experts, using the best available, up-to-date technology. With these three elements working together, [KISTI](#) believes it is able to achieve R&D success.

NTIS is a system where national agencies work together with world agencies to coordinate and manage research projects. The aim of NTIS is to support national science renovation through the construction of a comprehensive, scientific and technical information network of world-agencies at the national level. Major initiatives of NTIS are establishing a national, comprehensive scientific and technical information network for systemization and efficiency improvement of planning; managing and supporting R&D implementations; and managing performance information, etc. in R&D schemes through the whole business life cycle. This is a very ambitious, long-term project. The years 2005-2006 are the foundation-building stage, in which NTIS is establishing the ISP and studying the standardization technology. The system will be enlarged and adopted through 2007 and 2008. A comprehensive system should be in place by 2010.

[KISTI](#) is home to a national supercomputing center and has led national e-science and Grid initiatives. Emerging distributed collaborative scientific enterprises require desktop access to very large data collections, very large-scale computing resources, and high performance visualization. The national e-science environment of Korea will develop, implement, and deploy applications to utilize and create e-science technologies, infrastructure, and services. An essential feature of many of these services is that they will be used collaboratively, and by geographically distributed researchers in fields such as engineering, physics, earth science, bio-science, and chemistry. e-Science is a new paradigm of R&D activities which enables research productivity to be improved by networking diverse research resources. [KISTI](#) provides an infrastructure that enables flexible, secure, coordinated resource-sharing among dynamic collections of individuals, institutions and resources, to meet the needs of the domestic researchers.

The supercomputing infrastructure enables scientific discovery across a broad spectrum of science and technology, by the deployment and operation of a world class, persistent, production quality system. This becomes possible through a deep integration of existing national high-end

platforms, tightly coupled by a dedicated network and supported by innovative system and Grid software. [KISTI](#) provides a high-tech R&D infrastructure, composed of supercomputers and a high-speed research network, to researchers at domestic universities, research institutes, industrial entities and government agencies. It supports related utilization technology, and promotes application technology.

The supercomputing center is a resource center of secure world-class supercomputing systems; a technology center, helping Korean research communities to be equipped with a proper knowledge of supercomputing; a test-bed center, validating newly emerging concepts, ideas, tools and systems; and a value-added center, making best use of what the center has, in order to create new value.

KISTI's global R&D network includes GLORIAD and Korea Research Environment Open Network (KREONET). GLORIAD provides scientists around the globe with advanced networking tools that improve communications and data exchange, enabling active, daily collaboration on common problems. With GLORIAD, the scientific community can move unprecedented volumes of valuable data effortlessly, stream video and communicate through quality audio- and video-conferencing. GLORIAD exists today due to the shared commitment of the United States, Russia, China, Canada, The Netherlands and Korea to promote increased engagement and cooperation between their countries, beginning with their scientists, educators and young people. The benefits of this advanced network are shared with S&E communities throughout Europe, Asia and the Americas.

At present, Korea is actively carrying out Grid-based supercomputing and e-science collaborative researches throughout the advanced network infrastructure. These actions will be extended to GLORIAD activity so that they can play a major role in advancing the science and research development of the world. The candidate applications based on GLORIAD are expected to include high energy physics, ITER/fusion energy, atmospheric science, nanomaterials, bioinformatics and bioengineering, space sciences, and telemedicine applications, all of which require high bandwidth and quality for their scientific and technological researches.

KREONET is a national R&D network, run by [KISTI](#) and supported by the Korean government, since 1988. For science and technology information exchange and supercomputing related collaboration, KREONET provides Grid and e-science based, high-end, empowering application services with its global scale high performance network engineering. Currently over 200 organizations are member institutes of KREONET, including 50 government research institutes, 72 universities, and 15 industrial research laboratories.

In order to enlarge and secure all the services outlined in this talk, [KISTI](#) has formal links with many of the world's leading institutes:

- The National Center for Supercomputing Application (NCSA)
- Canadian Institute for Scientific & Technical Information (CISTI)
- Japan Science & Technology Corporation (JST)
- Institute of Scientific & Technical Information of China (ISTIC)
- Computer Network Information Center (CNIC)
- Library of Chinese Academy of Science (LCAS)
- Science & Technology Policy Research and Information Center (STPI)
- San Diego Supercomputer Center (SDSC)
- Institute of High Performance Computing (IHPC)
- National Center for Scientific and Technological Information (NACESTI)
- High Performance Computing Center Stuttgart (HLRS)
- All-Russian Institute for Scientific and Technical Information (VINITI)

[KISTI](#) has been securing a cooperative international network by exchanging information, technical specialists and technology resources and by holding joint seminars. KOSEN is a global network of Korean scientists and engineers who live in Korea and in foreign countries. This is highly regarded as a successful model of information sharing between scientists. There are 35,000 members from 40 different countries. Sixty-five percent of them have a master's degree or above. Information activities of KOSEN are information exchange on demand, information collection and analysis, online and off line community activities, and on-line consulting between members.

[KISTI](#) faces two challenges in securing all its functions: guaranteeing the connection of foreign networks, and laying strong foundations for cyber R&D environments. [KISTI](#) has been driving the Eastern North Asia Information Belt programs connecting the national information agencies of Korea, China and Japan, driving the cooperative development of scientific and technical information between the three nations and the connecting services; and implementing joint projects, such as information analyses, including information services and policies studies. [KISTI](#) is also trying to establish high-tech information infrastructure networks connecting R&D resources such as supercomputers and research networks, comprehensively, driving the joint use of high-tech research projects and equipment after establishing the Grid.

[KISTI](#) is proactive in establishing foundations for future high-tech renovated environments. Its systems embody the next generation, semantic information distribution environment, and digital archiving systems have been established. It is developing and applying high-tech analysis techniques for forecasts and evaluations of future R&BD information. It is expanding national e-science content, developing application technology and adopting ubiquitous application technology. As the institute which was assigned to play a pivotal role in establishing the national scientific and technical knowledge infrastructure, [KISTI](#) has been carrying its mission by constructing the Eastern North Asia R&BD network and firmly laying foundations for the cyber R&D environment.

The British Library ([BL](#))

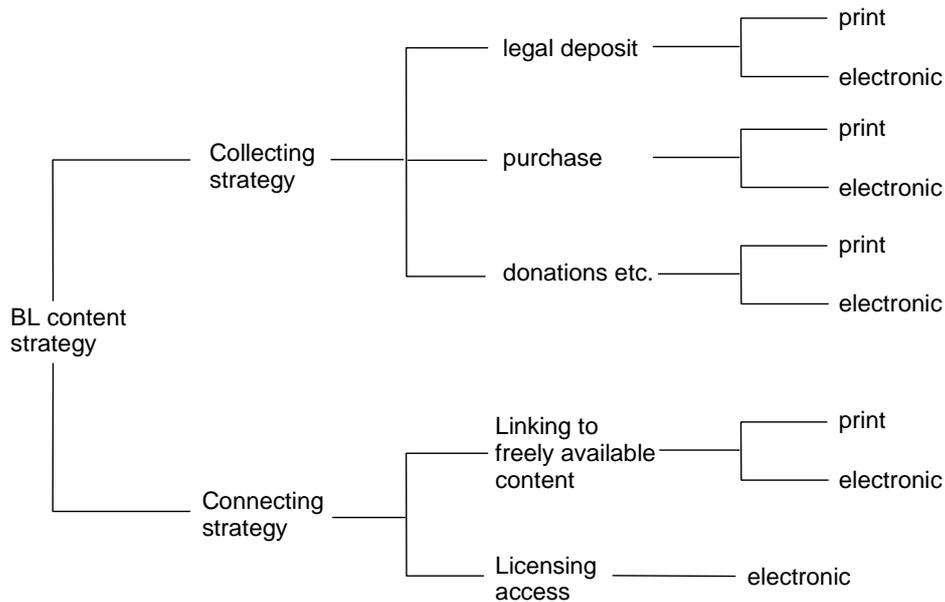
[Elizabeth Newbold, BL](#)

Newbold spoke about developing a content strategy at [BL](#). A strategy document for discussion was produced in 2005. [BL](#) has an acquisitions budget of about £15 million per annum but it still cannot be comprehensive in its acquisitions. An integrated library system was launched in 2004 that allows a more unified view of the collection for the first time. Previously, descriptions of the collections were largely geographically based or format-based. Now a disciplinary view is possible. The first focus is on purchased, not deposited material. ([BL](#) is a depository library in the United Kingdom.) Initially, [BL](#) is looking at what to purchase and why. Later, there will be an increasing focus on linking and connectivity: hence the interest in partnerships.

The strategy was developed over one year. Electronic Publishing Services Ltd. was engaged to input trend data. The project was overseen by an internal [BL](#) steering group, with guidance from the British Library Board and the British Library Advisory Council. There was also input from a specially convened external advisory panel with representatives from various disciplines and from the national and international library community.

A consultation paper was produced. This is a two-stage process. The first (the current stage) concerns overarching principles. The second stage will involve a detailed review of STM. The consultation paper has two parts: the main text and the appendices. Fourteen questions are distributed throughout the document. The whole was mailed to key stakeholders and put on the intranet. A twelve week period, April 25 to July 21, has been selected for feedback. Responses are coming in. [BL](#) targeted organizations that represent researchers and libraries but the consultation is open to all.

Newbold showed a diagram of the path from collection development to content strategy.



Standards and quality are paramount. [BL](#) has to have *trusted* partners.

The content strategy for STM presents huge challenges. First there are the needs of researchers. Research output continues to grow by 3-4% a year. There is a demand for rapid dissemination. The focus used to be on the scholarly journal. This is a more digitally advanced sector than many. Data is becoming increasingly important. Next, there are changes in scholarly communications and STM information. There are new business models such as author-pays. Some funding authorities are trying to make material more freely available. Open access repositories are appearing. Above all, digital preservation is essential.

[BL](#) has a high level content strategy and preliminary thinking. It continues to collect across all disciplines, with an increasing proportion in electronic form. It is reviewing its collecting profile across different areas of the world. Newbold noted that other speakers had also mentioned “the China challenge”. It is also developing a connecting strategy. As a research library, [BL](#) has never looked at data; now it must do so. Two documents are available on the Web: [Redefining the Library: the BL's Strategy 2005-2008](#), and [BL's Content Strategy: Meeting the Knowledge Needs of the Nation](#).