



Inforum 99

Sponsored by the Department of Energy

Digital Scientific and Technical Information Initiatives in an Interagency Context

Bonnie C. Carroll
President

Information International Associates, Inc.
Oak Ridge, Tennessee

May 6, 1999



New technologies that enhance the ability to create and understand information have always led to dramatic changes in civilization. The printing process unleashed the forces that led to the birth of the modern nation state. ...Now come distributed networks connecting a myriad of computers, ranging from megaflop machines and workstations to desktop and personal laptop units. ...There is no longer any doubt that such machines will reshape human civilization even more quickly and more thoroughly than did the printing press.”

**Albert Gore, *Scientific American*, Sept. 1991
“Infrastructure on the Global Village”**





Visions of the Future

- Apple's Information Navigator
- Star Trek



In the Paleo-electronic World, Predator/Prey Relationships are Rapidly Reversing.

- Print world talks of “food for thought”
- Networked world users will be the prey
and information is the predator

*-- Paul Peters, Talk at NASA Third Annual Foreign
Acquisitions Workshop, September 23, 1993*





Objectives of this Talk

- **Share a common vision**
 - “From Institutionalized Anarchy to Federated Networks”
- **Provide a perspective on other/inter-agencies**
- **Make connections and conversation**
- **Move the agenda forward**



Why Here?

- **DOE is an active member of CENDI**
 - Leadership role in the evolving paradigm of STI management
- **Committed to building a cooperative information infrastructure in STI**
- **CENDI context**





Why Me?

- Secretariat Director of CENDI
 - Biodiversity and Ecosystems Informatics Working Group (NSTC)
- Consultants in STI planning and management
- Tenure -- three decades of exposure



What is CENDI?

- An interagency group of senior STI managers who have agreed to cooperate by
 - Exchanging information and ideas
 - Sharing resources
 - Undertaking joint initiatives
 - From policy to standards to operations
- Agencies manage over 90% of the \$ 80B federal R&D budget

www.dtic.mil/cendi/





CENDI's Vision

Federal STI agencies will have a cooperative *enterprise* where capabilities are shared and challenges are faced together so that the sum of accomplishments is greater than each individual agency can achieve on its own.



CENDI's Mission

CENDI's mission is to help improve the productivity of federal science- and technology-based programs through effective scientific, technical, and related information support systems.





CENDI Goals

- ◆ **Goal 1. *Coordination and leadership:*** Provide coordination and leadership for information exchange on important STI issues.
- ◆ **Goal 2. *Joint projects:*** Undertake joint projects by two or more CENDI agencies to increase effectiveness and use of STI systems.
- ◆ **Goal 3. *Education:*** Promote an understanding of STI and STI management



CENDI Membership

CENDI is an interagency executive group composed of Senior STI Managers

- 10 major programs
- 9 U.S. federal agencies
- 7 Cabinet departments
- 2 Independent agencies

- COMMERCE - National Technical Information Service (NTIS)
- ENERGY - Office of Scientific and Technical Information (OSTI)
- EPA - Science Information Management Coordination Board (SIMCorB)
- NASA - Scientific and Technical Information Program (STI)
- NATIONAL Libraries of Agriculture, Education and Medicine (NAL, NLE, and NLM)
- DEFENSE - Defense Technical Information Center (DTIC)
National Air Intelligence Center (NAIC)
- INTERIOR - USGS Biological Resources Division (BRD)





CENDI is Formally Established Under an Interagency MOU

- 1973 □ Phase-out of COSATI
- Mid-1970's to Mid-1980's □ Cooperation on ad hoc basis
- 1982 □ Regular but informal meetings
- 1985 □ Formal interagency MOU among four agencies (NTIS, DOE, DoD, and NASA)
- 1987 □ National Library of Medicine (NLM) joined
- 1988 □ Secretariat created
- 1993 □ National Air Intelligence Center (NAIC) joined
- 1995 □ DoI/USGS, Biological Resources Division (BRD) joined
- 1998 □ National Libraries of Education & Agriculture (NLE, NAL) joined
- 1999 □ Environmental Protection Agency (EPA) joined



Historical Perspective

**By the end of the 1960's, we had
evolved an effective STI
Infrastructure.**





The Traditional Science and Technology Information Infrastructure (STII)

[Graphic not available electronically]

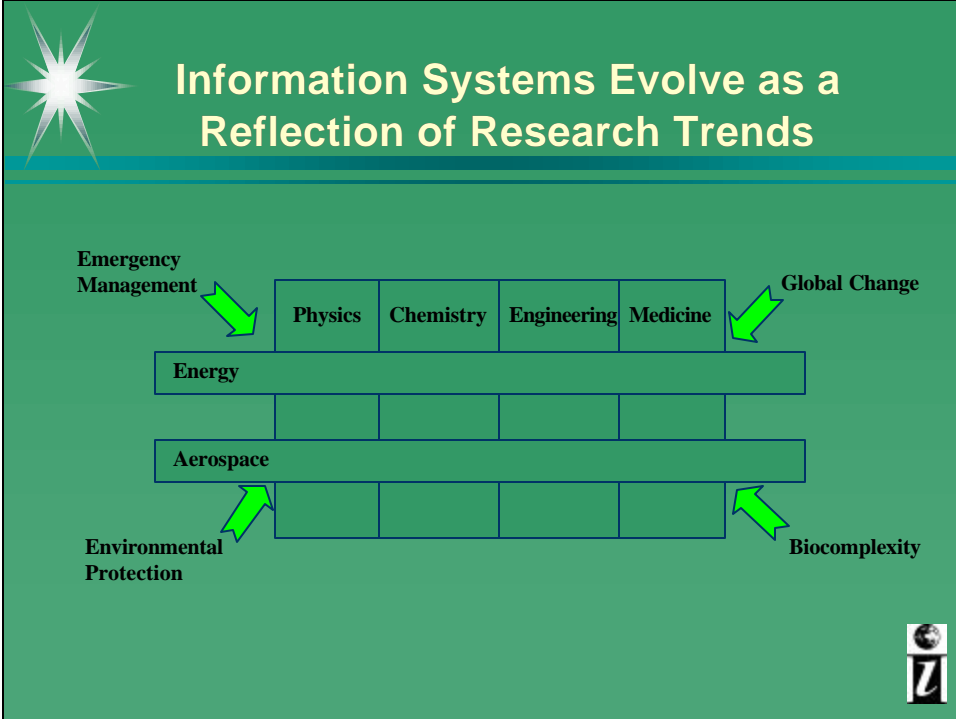


The Need for the Next Generation System

- Transition from information systems to decision support
- Forces are both technological push - user pull
- Move from passive to active systems - extend intellectual power
- Needs of our age are in crisis management and decision making
 - Need to integrate STI and societal information (STSI)

*1978 National Science Foundation Report,
"Passing the Threshold into the Information Age"*





- ## By the 1980's, the Theme was Dealing with the Crisis of Information Overload
- Irony of Information Age is not a problem of scarcity
 - Science (8/12/83) article shows that information availability is increasing more rapidly than the ability to absorb it
 - Due to effectiveness of electronic access systems, too much “relevant” information is identified
 - Need the right information at the right time
 - With technological advances, the problem became magnified
 - Algorithms for quality discriminations are not yet developed
-



The Need for the Next Generation System

- Need for interconnected national information technology network
- New information infrastructure
 - Stores information in forms accessible
- Provided means of communication
- Aids in information manipulation to create new insights and knowledge
- Provides human and computer assistant; i.e., “user friendly”

“Infory” of the Future:

- Distribute in physical sense
- Professional career structure to identify and respect
- Dissemination of reliable, personalized software

*1989 National Academy of Sciences Report,
“Information Technology and the Conduct of Research”*



By the 1990’s, the Paradigm in STI Began to Shift

- Basic changes in structure of science itself
- Increasing commercialization and marketability of science
- Advances in technology
- Growth in volume of information
- Lack of coordinated information policy and federal STI leadership





Question of the Month on Internet (1993):

In what month will the number of bytes run through the Internet exceed the number of atoms in the Universe?



Basic Changes in Structure of Science Itself ⁽¹⁾

- **Disciplinary, mission, problem-solving orientations**
- **Blurring boundaries among science, social science, politics, marketing, ethics**
- **Specialization but linkages**
- **New tools and new methods -- computational sciences**





Basic Changes in Structure of Science Itself (2)

- Big science, international science
- Increase in commercialization and marketability of science
- Shift from military to industrial R&D
 - Global politics
- Changing user expectations
 - Computer literacy of next generation
 - User community initiatives



By 1999, We're Reinventing our Roles

- Agency STI programs redefining missions
- Budget allocations are problematic
- Publishers are questioning primary/secondary relationships
- Authors are experimenting
- User expectations are changing
- Madison Avenue/Wall Street are "IN"



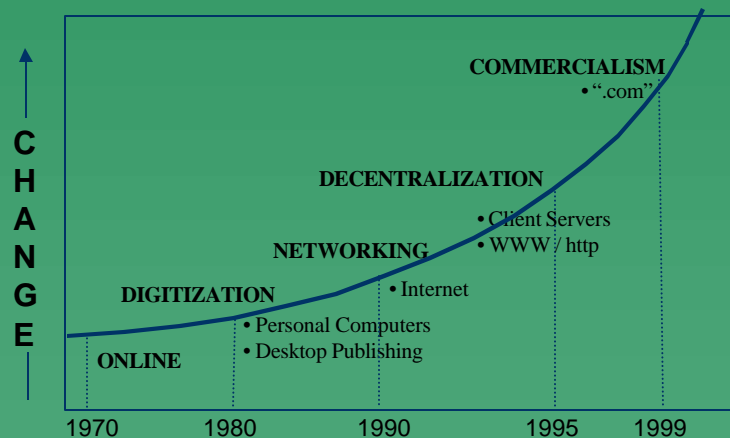


Visions for Content-based Next Generation Information Infrastructures are Appearing at High Levels

- PCAST, 1997
- *“Teaming with Life: Investing in Science to Understand and Use America’s Living Capital”*



TECHNOLOGIES OVER TIME





CENDI Panning Themes Set the Stage

- 1990 □ The Value of Cooperation: Positioning for the 1990's
- 1991 □ Building Strategic Alliance
- 1992 □ Critical Connections: Cooperation to Enhance Information Value and Use
- 1993 □ Navigating in a Networked World
- 1994 □ Building a Digital Information Network
- 1995 □ Changing Missions -- Changing Technologies
- 1996 □ Opportunity in Change
- 1997 □ Changing STI Management in a Networked Environment
- 1998 □ Enterprise Networking and STI Futures
- 1999 □ The Virtual STI Enterprise: Completing the Connections



CENDI STI Definition

STI is information that derives from ... research, development, and deployment (RD&D) results of the efforts of scientists and engineers and individuals supporting their work. STI includes new theory and information obtained from experimentation, observation, instrumentation, or computation in the form of text, numeric data, or images. STI may be further transformed, described, evaluated, synthesized, and recorded in print, micrographic, magnetic, optical, or other media to enhance its communication and its usefulness and value to a wide spectrum of users and uses.





“The Changing R&D Information Economy in the Digital Age”

-- Robert Ubell, September 1997



As the Digital Revolution Progresses, Both Costs and Benefits will Rise

Complexity in the background

- Number of scientists, engineers, physicians and other technical personnel working today = total who existed since ancient Greece
- Information explosion increasing
 - 165,000 periodicals published, 2000 added annually (10,000 peer reviewed)
- Publishers have yet to convert entirely to electronic delivery
- Dramatic differences among the scientists
 - Some scientists still rely on traditional tools
- Prices of S&T journals increased far in excess of inflation





Obstacles to Electronic Utopia

- **Publishing**
 - First-copy costs do not disappear -- costs/circulation spiral
- **Technology**
 - Installing and maintaining technological infrastructure
- **Human Resources**
 - Trained technical staff with new roles and responsibilities is more expensive for electronic processes
- **User Expectations**
 - Technology generates greater demand, increases user expectations, customization
- **Transition**
 - Accommodate costly “mixed economy”



Key Judgements

- Fully electronic information management is in its infancy and the full life cycle cost is not yet understood
- The real payoff of the digital world may not be simply economic, but, instead, the ability to use better information for better decision-making
- There is often a shift in where costs are incurred and savings are gained. The information management function may increase costs but offsets to users provide an overall life cycle return on investment for the agency or nation.





**CENDI FY99 Planning Session
The Virtual STI Enterprise: Completing the
Connections**

- **NSF - DL-2: Lesk (Information and Intelligence Systems) interest in science crosscut**
- **CENDI DL Study**
- **National Library for Science & Technology (NLST) White Paper**
 - To explore concepts



***“CENDI Agency Digital Library
Initiatives: Toward a Digital Future”***

-- Gail Hodge, August 1998





GENERAL FINDINGS

- **Agencies differ but there is considerable overlap**
 - Stage of digital-virtual library development
 - Primary audience
 - Material in the collections
 - Components of a digital-virtual library that were emphasized
- **Agencies agreed ~**
 - Overlap in research interests
 - Opportunities to collaborate



DESCRIPTIONS OF AGENCY PROJECTS

- **DOE - EnergyFiles**
 - Virtual library with contributions from OSTI, other DOE offices and programs, academia and industry
 - InfoBridge
 - OSTI full text of technical reports which is part of EnergyFiles
- **NASA - NASA Electronic Library**
 - Will seek to provide a single point of access (virtual library) to distributed sources on other NASA web pages
- **NLM - PubMed, History of Medicine, etc.**
 - Numerous digital projects available from their web page; emphasis on PubMed with connections to full text from the publishers
- **USGS/BRD - National Biological Information Infrastructure**
 - Collaborative project emphasizes data sets, BRD publications, geospatial referencing and future modeling capabilities





DESCRIPTIONS OF AGENCY PROJECTS, *continued*

- **NAL - Agricultural Network of Information Centers**
 - Expertise from collaborators in academia; each center has a different expertise
- **DTIC - Defense Virtual Library**
 - Emphasizes non-textual objects; working with CNRI and other collaborators on the distributed Repository Architecture
- **NAIC - Digital Library Input System and Portal**
 - Full-text input/tagging system for English and foreign material (DLIPS); Portal can include multiple libraries in addition to CIRC (combination digital-virtual)
- **NLE - General Educational Material (GEM)**
 - Lesson plans and teachers of excellence; AskA digital reference service
- **NTIS - SpecFinder and other projects under FedWorld**
 - Provides DoD solicitations linked to the full text of the referenced standards; supported by the NTIS database and ordering system



DIGITAL-VIRTUAL LIBRARY COMPONENTS

- **Collection development**
- **Indexing/categorization, including vocabulary development**
- **Metadata and standards**
- **Search engines**
- **Profiling/user push technologies**
- **Archiving and preservation**
- **Digitizing technologies and the workflow to support them**
- **Personnel and cultural issues**





COLLECTION DEVELOPMENT

- Tension between focusing on the defined subject scope or the specific needs of an audience as it changes
- Different formats and document types, multimedia, numeric data, audio, video, etc.
- Several agencies planning to connect to electronic journals or to publisher web sites (a la PubMed)
- Looking to collaborators to provide material to supplement that of the agency



INDEXING/CATEGORIZATION

- Using subject categories or controlled keywords for:
 - More precise searching of metadata for non-textual objects
 - Express themes for web pages
- Agencies have utilized or modified their subject categorization schemes and thesauri
- Other agencies are developing new vocabularies





METADATA AND METADATA STANDARDS

- Numerous standards are being used (FGDC, DoD GILS, Dublin Core, etc.), some based on requirements of the community of which the agency is a part (*see report on CENDI Metadata Initiatives*)
- Extensions are viewed as important to handle unique data and legacy records
- Interoperability is a concern (some agencies are “watching” RDF and XML)



SEARCH ENGINES

- Variety of search engines are being used
 - Entrez (NLM)
 - InQuery and Blue Angel (DTIC)
 - Verity (NTIS)
 - RetrievalWare (NAIC)
 - OpenText (DOE)
 - ISite (BRD)
 - UltraSeek (NLE)
 - (NASA)
 - (NAL)
- Some agencies use Z39.50 to provide distributed searching
- Distributed search engines are of great importance and one of the most important research interests





PROFILING AND PUSH TECHNOLOGIES

- Replaces previous SDI services - qualitatively the same but new technology
- Way to provide different community views (public, professionals, technicians, etc.)
- How to present the “pushed” information; e-mail is most prevalent approach, some discussion of personal web pages



ARCHIVING AND PRESERVATION

- NAL’s report on digital preservation has been released
- Other agencies have considered issues of archiving; most are taking on this role if they “own” the data
- Problem arises when dealing with a virtual library system outside their span of agency control
- Lack of URL persistence - “handle” system and Digital Object Identifier are being planned





DIGITIZING TECHNOLOGIES AND WORKFLOW

- Different scanning and OCR technologies; some are being upgraded
- NAIC is addressing workflow issues in major reengineering
- Transitional flows are currently in use:
 - Scanning and OCR of hardcopy
 - Electronic in various formats
 - Pointing to full text on distributed servers



CULTURAL AND BUSINESS ISSUES

- New technologies and work flows have raised personnel and training issues
- Collaboration with other companies raises:
 - Cultural issues
 - Intellectual property and copyright issues
 - Who is the audience, how to provide customer support
- Users expectations and realities
- Economics -- what are the real lifecycle costs?
 - Who will pay for the digital-virtual library?





POSSIBLE RESEARCH AREAS

- Metadata interoperability
- Distributed searching
- Distributed vocabularies that can be integrated
- Tools for profiling and customizing collections
- Usability and evaluation methodologies/metrics
- Culture of digital libraries (among collaborators, users, and agency staff)
- Economics and funding models for digital-virtual libraries
- Some of the agencies are collaborating on DL-2 and KDI research proposals
 - Some providing sponsorship



Toward a National Library of Science and Technology: Building on the Present -- Creating the Digital Future

- White paper ~ October 1998

- Premise:

We have a unique opportunity to embrace the Information Age Technology to capitalize on the national STI resource.

-- Bonnie C. Carroll

-- Gail M. Hodge





Building Blocks

- **National Library Precedents ~ NLM, NAL, NLE**
 - Missions have a public good aspect and addressed a national priority
 - An information resource existed in a mission agency and its value was recognized as transcending its sponsor.
- **NL concept in digital networked world**
 - Theme-based information infrastructures
 - NBII
 - EnergyFiles: VLEST
 - DVL
 - PubMed
 - AgNIC
 - All-Source Digital Library



Completing the Connections

- **Federated network based on common standards and mutual self-interest**
 - Built on strength of existing resources
 - No “Edifice Complex”
- **Cutting across disciplines and missions**
 - Allowing better customization
- **CENDI was well-positioned to promote cooperation**





How Does CENDI Move the Agenda Forward?

- “Think globally; act locally”
- Top down vision - August 1999
- Information exchanges on agency initiatives
- Three specific activities
 1. Undertake a pilot project: ILSE
 2. Address specific issues: DEA
 3. Continue development of standards, tools, technologies: Vocabularies



ILSE: CENDI AGREED TO PILOT INITIATIVE

- Using Intelligence Community Technology: “Broadsword”
 - Universal logon
 - Cross database searching using “plug-ins” to map to common definitions
 - data elements
 - search commands
 - parameters
 - syntax
 - Three agencies: DTIC, NAIC, DOE
 - Demonstration at GovTechNet '99, June 15, 1999





Study Digital Electronic Archiving ICSTI / CENDI Cooperation

- Identified as key concern
 - The exponential growth in the creation and dissemination of digital objects has emphasized the speed and ease of short-term dissemination with little regard of the long-term preservation of digital information
- Digital information is inherently more fragile than traditional technologies
 - It is more easily corrupted or altered without recognition
 - Because of technology, the time frame to consider archiving is shorter
- Those who did not consider themselves archivists now drawn in either
 - infrastructure
 - intellectual property issues
 - user expectations



Study Parameters

- Identify state-of-the-art and -practice
 - STI
 - International
 - All media
- 50 projects identified
 - Qualified 18
 - 7 countries
 - All sectors
- Models
- Life cycle managers/roles
- Best practices
- Issues
 - Costs/resources
- Recommendations
- Presentations to ICSTI, CENDI, World Science Congress





Hold a Workshop on the Role of Thesauri and Controlled Vocabulary in Internet-hosted Databases

- Fundamental concerns in knowledge organization
- Number of vocabulary-based initiatives
 - UMLS
 - NBII/CERES
 - ICSTI
- Key to crossing disciplines and missions
- Fall 1999
- Subject analysis and retrieval, cataloging, user education working groups



Why You?

Can We Build Toward a Shared Vision?

[Camel Graphic not available electronically]





Changing User Expectations

**1/3 of the population know what
the Internet is**

2/3 think it is a good idea

