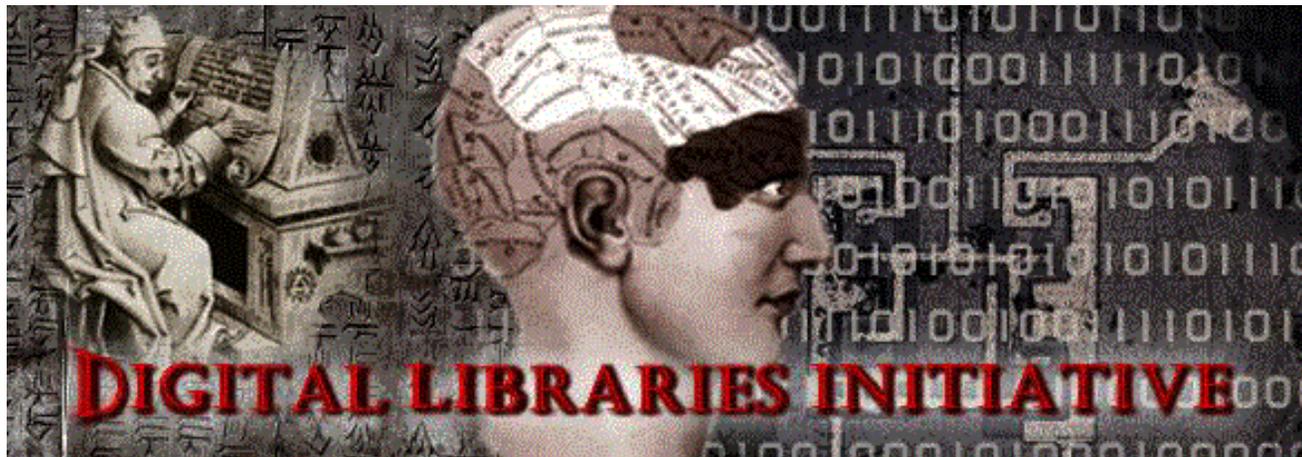


# **DIGITAL LIBRARIES INITIATIVE**

## **An Interagency Program of Research and Applications**



*[www.dli2.nsf.gov](http://www.dli2.nsf.gov)*

Stephen M. Griffin  
National Science Foundation

# Digital Libraries Initiative (DLI) Program Profile (Phase 1)

---



- Six university-led projects; similar project model for each
- \$24M total over four years, ending fall 1998.  
(~\$1M/year per project)
- Each project required to:
  - Carry out fundamental research
  - Create a large testbed
  - Work with partners
  - Acquire substantial cost sharing (1:1 is norm)
  - Demonstrate leadership for larger community
- Cooperative agreements, not grants
- All-Project meetings every six months
- D-Lib Magazine (DARPA sponsored)

# Digital Libraries Initiative (DLI) Phase 1 Projects

---



## Project/Research Focus

## research goal

### **Carnegie Mellon University: Digital Video Libraries**

- speech, image and natural language technologies integration



*full-content search and retrieval of video segments*

### **Univ of Michigan: Intelligent Agent Architectures**

- software agents; resource federation; artificial service market economies; educational impact



*new DL cross-disciplinary capabilities, intellectual perspectives and linkages*

### **Stanford Univ: Uniform Access**

- interoperability; protocols & standards; distributed object architectures; interface design for distributed information retrieval



*general access, extensibility for heterogeneous distributed resources*

# Digital Libraries Initiative (DLI) Phase 1 Projects

---



## Project/Research Focus

## research goal

### **Univ of California, Santa Barbara: Geographic Information Systems**

- spatially-indexed data; content-based retrieval; image-compression; metadata



*resources for geosciences research and education communities*

### **Univ of Illinois: Intelligent Search and the Net**

- large-scale information retrieval across knowledge domains; semantic search; SGML; user/usage studies



*semantic retrieval across the net; alternatives for publishers of scientific journals*

### **Univ of California, Berkeley: Media Integration and Access**

- new models of “documents”; natural language processing; content-based image retrieval; innovative interface design



*new models and services for multi-media information management in a networked world*

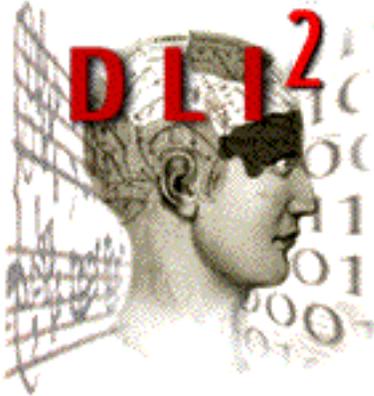
# DLI Collaboration and Partnering



*Flow of Resources, Technologies, Knowledge, Intellectual Products*



<p><b>Computer &amp; Communications Companies</b></p> <ul style="list-style-type: none"> <li>Digital Equipment Corp</li> <li>Xerox Corp</li> <li>Xerox PARC</li> <li>Intel Corp</li> <li>Apple Corporation</li> <li>Bellcore</li> <li>Eastman Kodak Co</li> <li>IBM</li> <li>Lockheed</li> <li>Interconnect Tech Corp</li> <li>Enterprise Integration (EIT)</li> <li>Bellcore</li> <li>Interval</li> <li>Microsoft Corp</li> <li>Bell Atlantic Network Services</li> <li>AT&amp;T</li> <li>Hewlett Packard</li> <li>United Technologies</li> <li>Softquad</li> <li>BRS/Dataware</li> <li>Spyglass</li> <li>Hitachi</li> </ul>	<p><b>Publishers/Content Providers</b></p> <ul style="list-style-type: none"> <li>Elsevier Science Group</li> <li>Encyclopedia Britannica</li> <li>McGraw-Hill Publishers</li> <li>Dialog Information Services</li> <li>O'Reilly</li> <li>WAIS Inc</li> <li>QED Communications</li> <li>John Wiley &amp; Sons</li> <li>U.S. News &amp; World Report</li> <li>M&amp;T Publishing</li> <li>Tribune Company</li> <li>UMI</li> </ul> <p><b>Professional Societies</b></p> <ul style="list-style-type: none"> <li>American Math Society (AMA)</li> <li>ACM</li> <li>IEEE</li> <li>American Institute of Aeronautics and Astronautics (AIAA)</li> <li>American Physical Society</li> <li>American Institute of Physics</li> <li>NCGIA</li> <li>Association of Research Libraries</li> </ul>	<p><b>Other Universities</b></p> <ul style="list-style-type: none"> <li>SUNY Buffalo</li> <li>Univ of Maine</li> <li>Univ of Arizona</li> <li>Open University, U.K.</li> <li>Univ of Wisconsin</li> <li>Univ of Colorado</li> <li>MIT</li> <li>Cornell Univ</li> </ul> <p><b>Libraries</b></p> <ul style="list-style-type: none"> <li>Project Site Univ Libs</li> <li>USGS Library</li> <li>Library of Congress</li> <li>California State Library</li> <li>Sonoma County Library</li> <li>St. Louis Public Library</li> <li>New York Public Libs</li> </ul> <p><b>International Orgs</b></p> <ul style="list-style-type: none"> <li>ERCIM</li> </ul>	<p><b>Primary &amp; Secondary Schools</b></p> <ul style="list-style-type: none"> <li>Project-local comm schools</li> <li>Fairfax County Public Schools</li> <li>Winchester-Thurston School</li> <li>Ann Arbor Public Schools</li> <li>Stuyvesant High School, NYC</li> <li>Shasta County Ofc of Edu</li> </ul> <p><b>Government Agencies and Labs</b></p> <ul style="list-style-type: none"> <li>DMA/CIO</li> <li>U S Navy</li> <li>USGS</li> <li>NASA/ARC</li> <li>Res Agcy of California</li> <li>San Diego Assn of Govts</li> </ul> <p><b>Other/Non-Profits</b></p> <ul style="list-style-type: none"> <li>CNRI</li> <li>Environmental Systems Res Inst</li> <li>Mellon Foundation</li> <li>Kellogg Foundation</li> <li>Getty Foundation</li> </ul>
---	---	--	--



# Digital Libraries Initiative – Phase 2

## Sponsoring Agencies and Partners

*[www.dli2.nsf.gov](http://www.dli2.nsf.gov)*

**NSF (CISE, SBE, EHR/DUE)**

**DARPA**

**National Library of Medicine**

**Library of Congress**

**National Endowment for the Humanities**

**NASA**

**FBI**

**National Archives**

**Smithsonian**

**Institute for Museum and Library Services**

# *Digital Libraries Initiative - Phase 2*

---



**Core Sponsors: NSF, DARPA, NLM, LoC, NASA, NEH**

**~\$8-10 million/yr for 4-5 years (beginning FY98)**

- **sponsor a full-spectrum of activities: fundamental research, content & collections development, domain applications, testbeds, operational environments, new resources for education and preserving America's cultural heritage**
- **address topics over entire DL lifecycle: information creation, dissemination, access, use, preservation, impact, contexts**
- **implement a modular, open program structure: add new sponsors, performers, projects at any time**

## **Program Goals:**

**new DL research, technologies and applications to advance the use of distributed, networked information of all types around the nation and the world**

# Comparison of DLI with DLI - Phase 2



## DLI

### research

broad, technology-centered

### testbeds

for technology research

### content/collections

donated to projects

### infrastructure

limited testbed development

### context

primarily user evaluation

1994

## DLI - Phase 2

### research

refined technical scope; extend to new areas and dimensions in the DL information lifecycle

### testbeds

for DL research with added emphasis on interoperability & technology integration

### content/collections

increased emphasis on content, collections development and management

### infrastructure

operational DLs with collections of value to domain and other “communities” of users

### context

understanding DLs in domain, economic, social, international contexts; DLs as HuCS

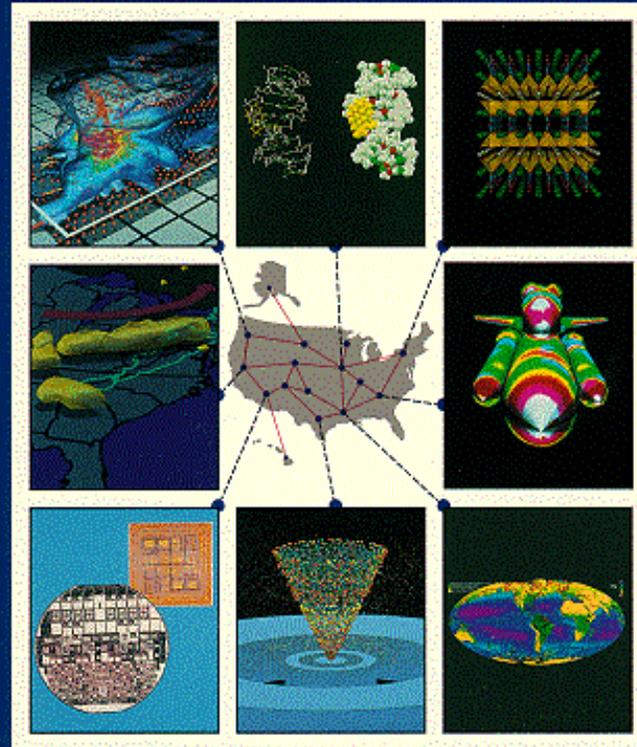
1998

# The Federal High Performance Computing and Communications Program

1992-1996

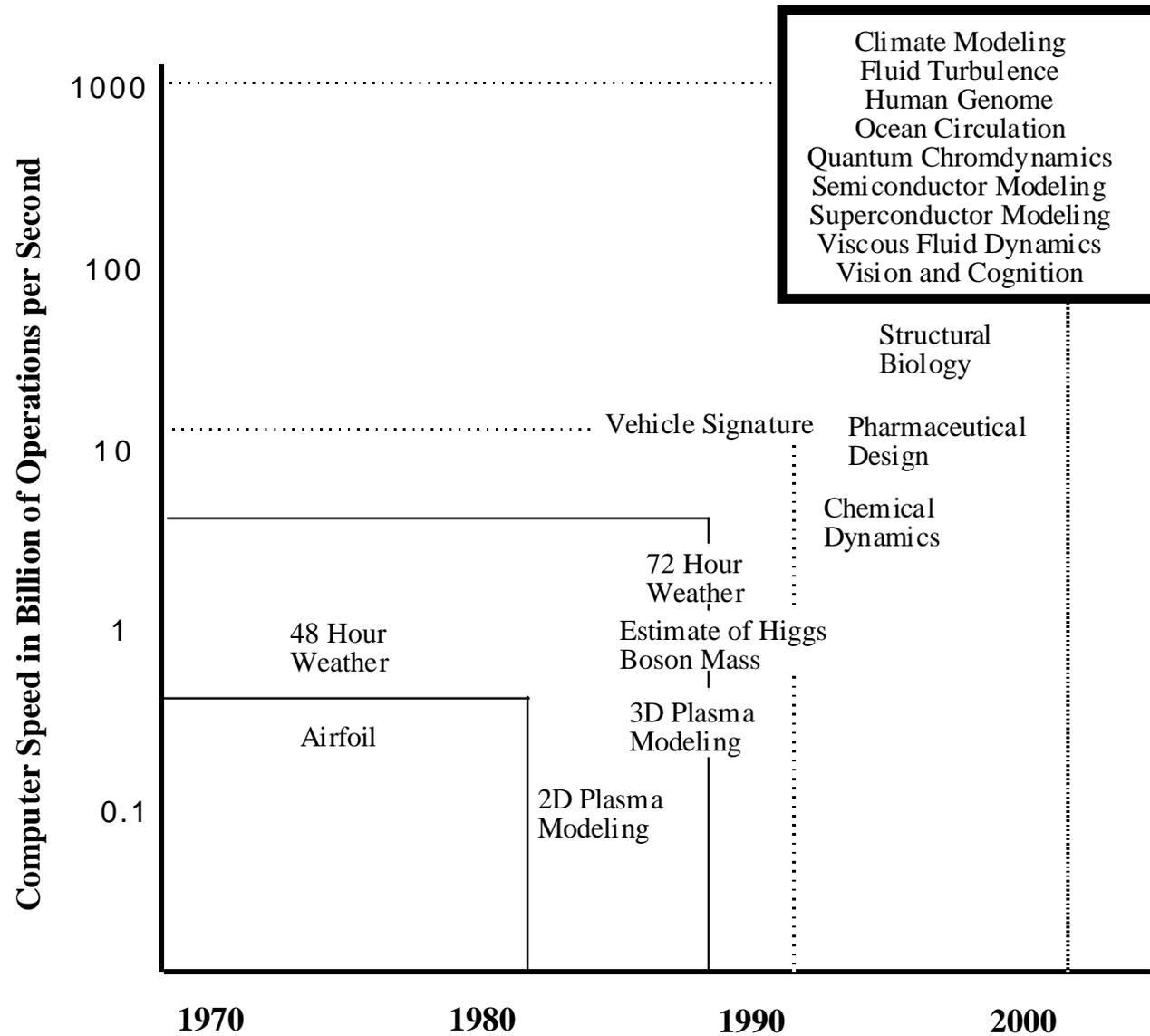
## Grand Challenges: High Performance Computing and Communications

The FY 1992 U.S. Research and Development Program

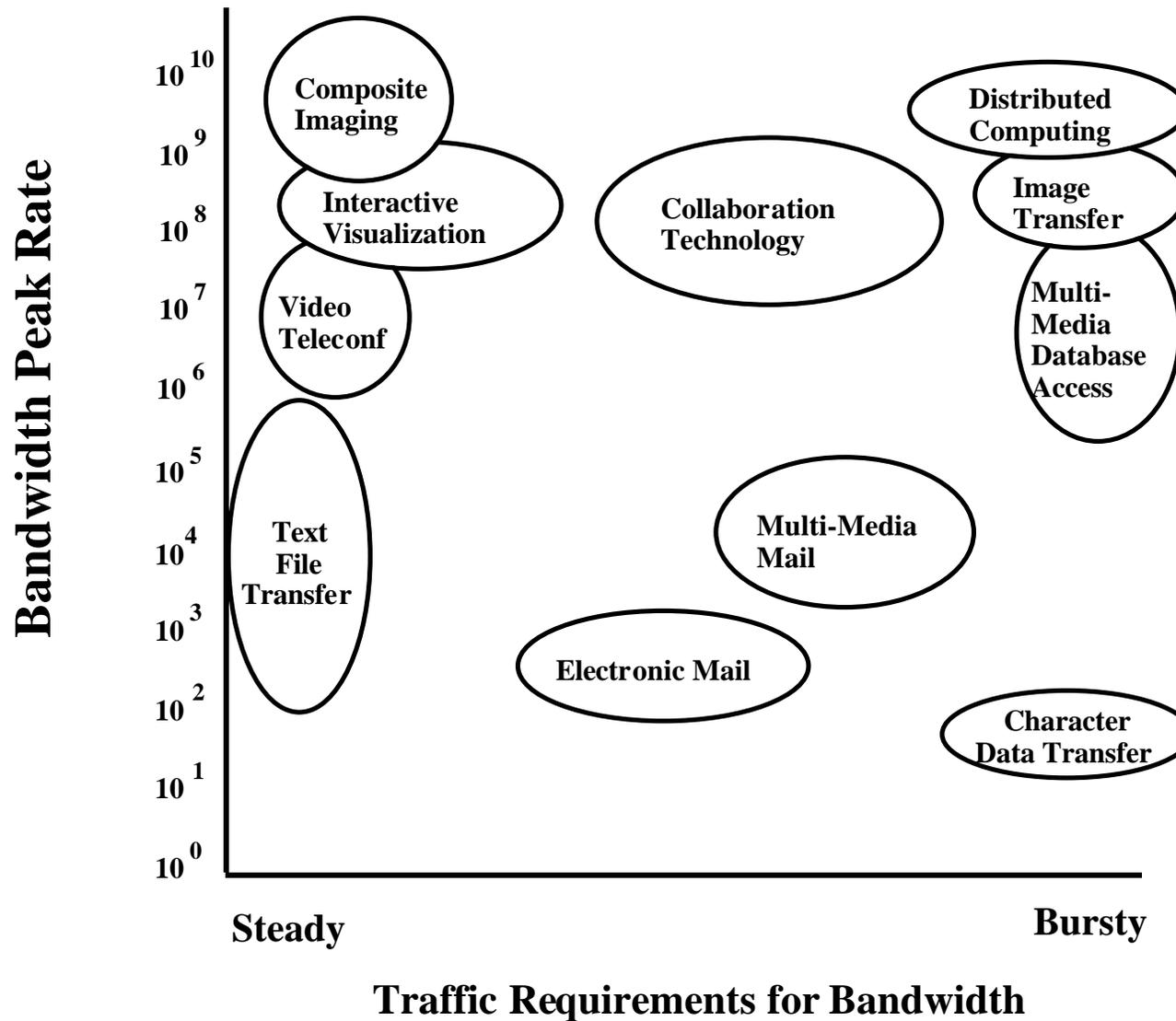


A Report by the Committee on  
Physical, Mathematical, and Engineering Sciences  
To Supplement the President's Fiscal Year 1992 Budget

# Grand Challenge Requirements



# NREN Applications by Bandwidth and Traffic Characteristics

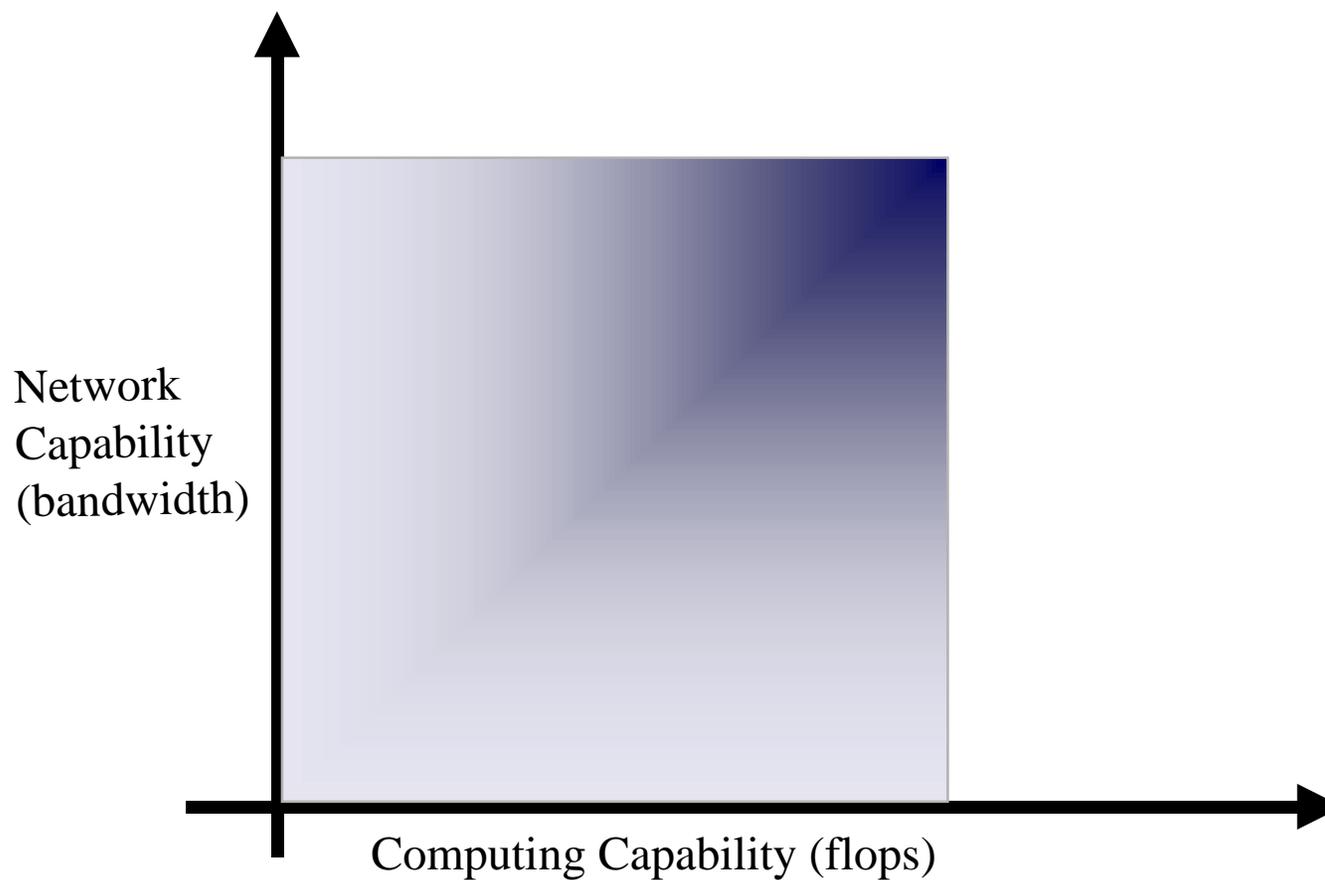




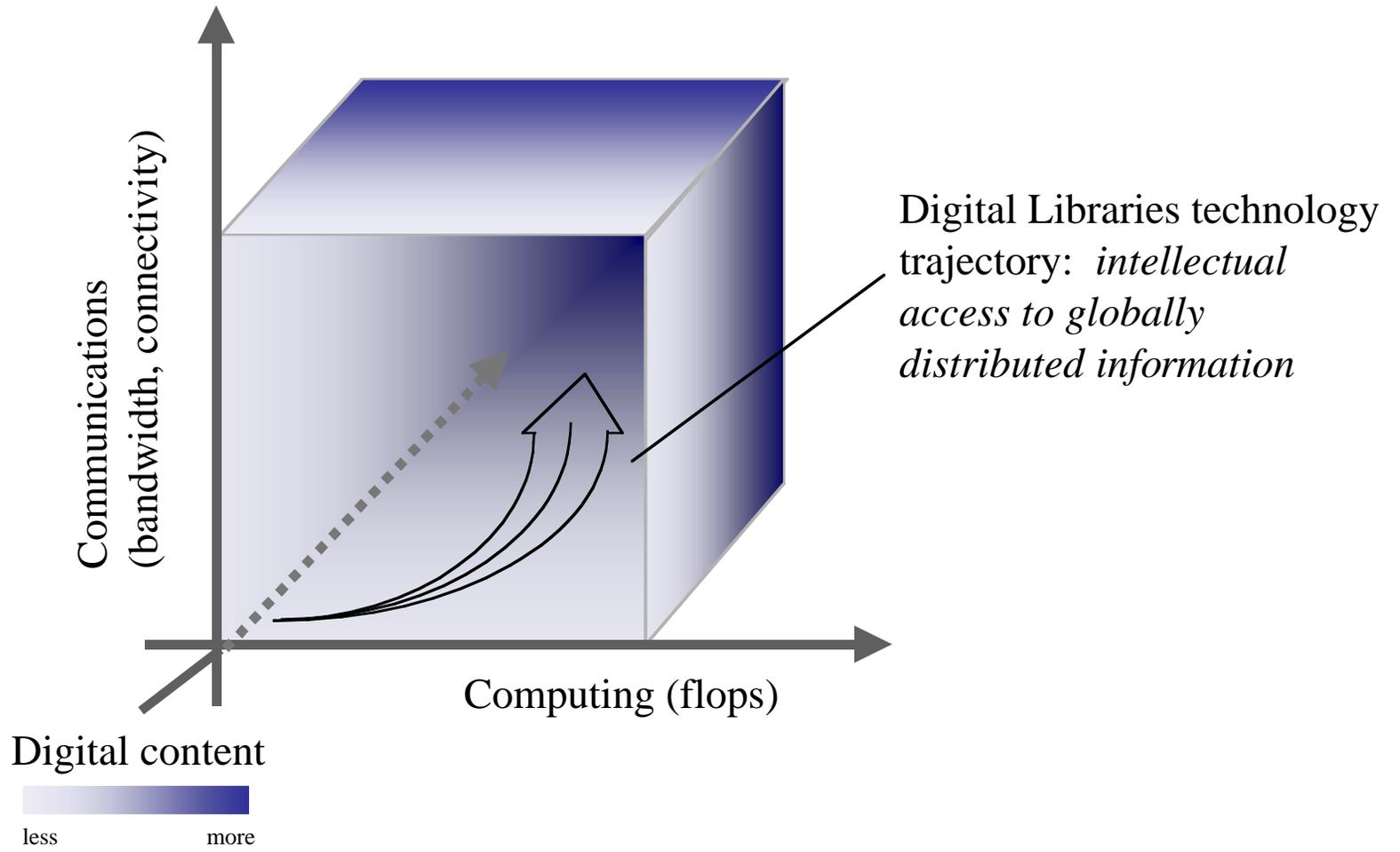
## *Two Dimensional Thinking of Early 1990s...*

---

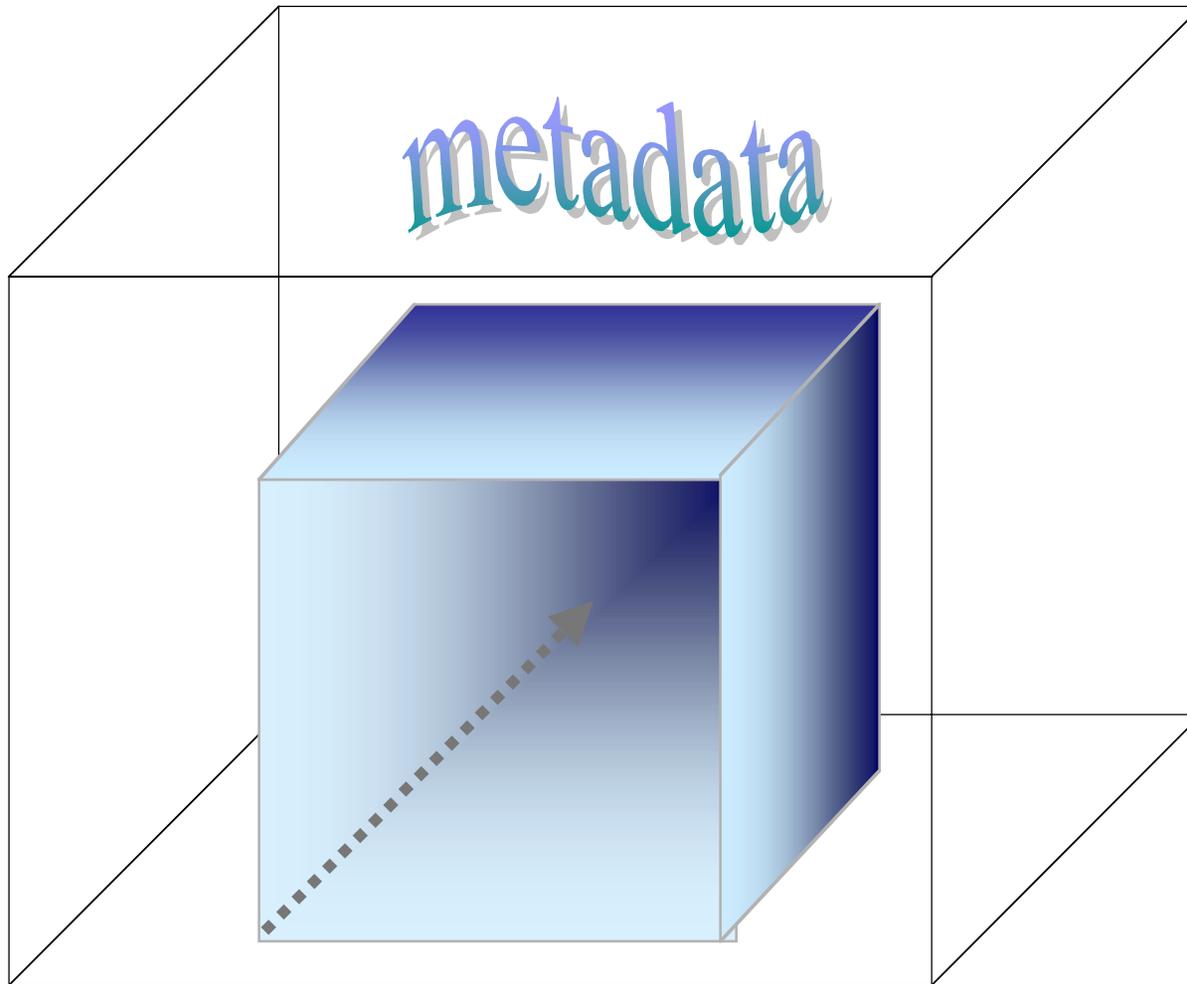
---



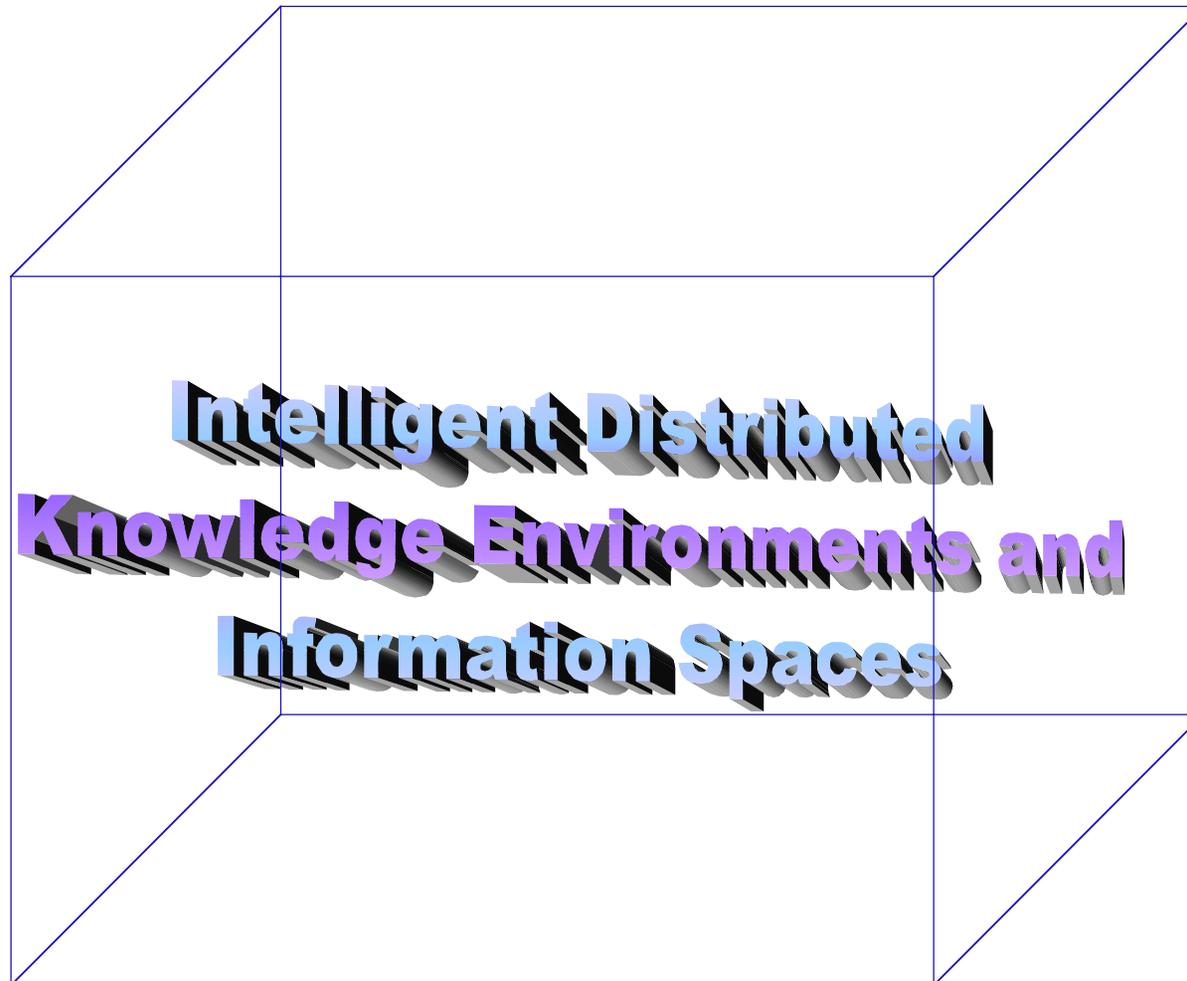
# Three Dimensional Thinking of mid-90s...



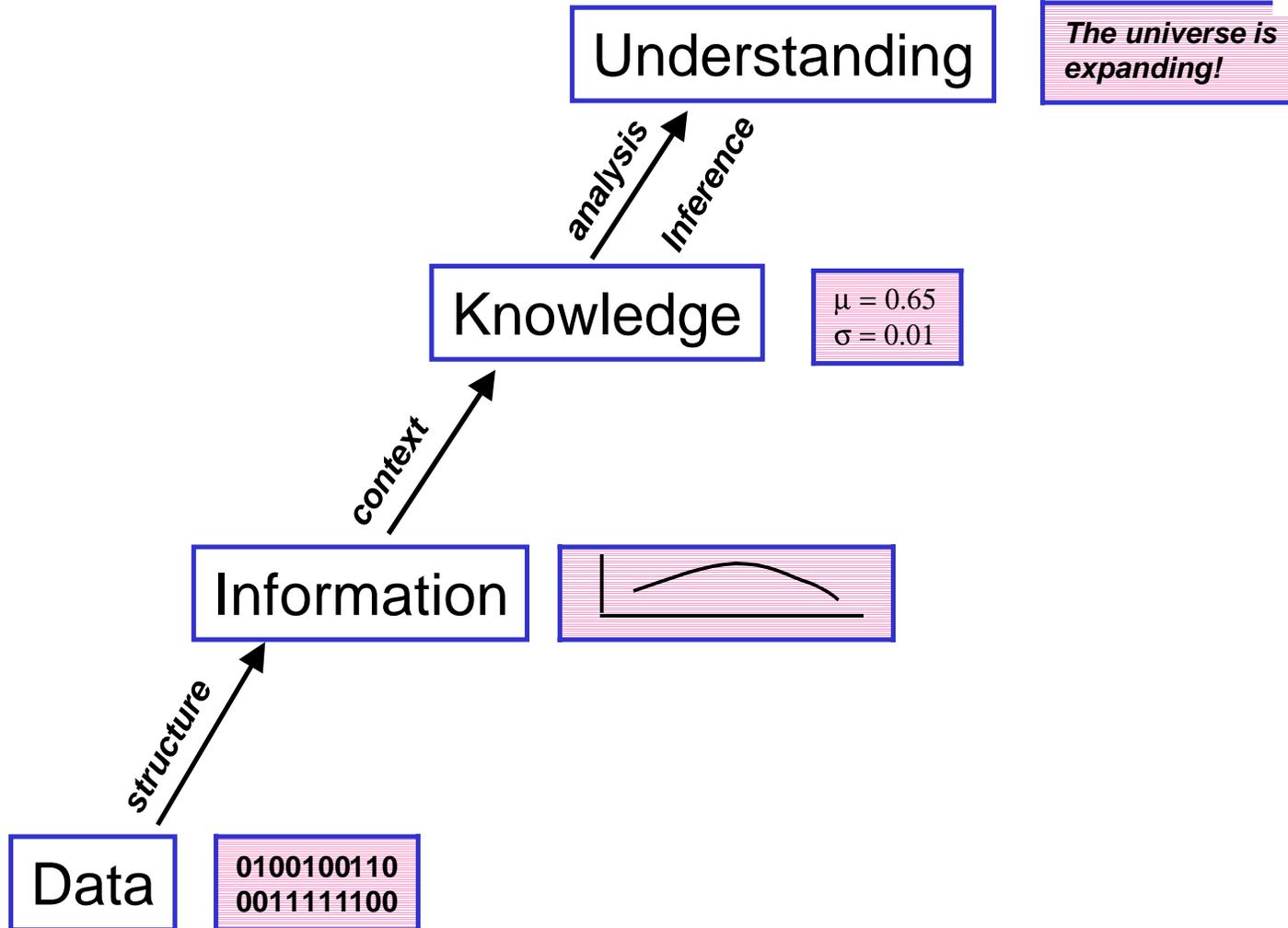
*Today: Emphasis on Context and Structure*



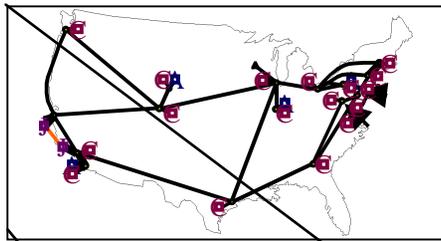
*Next: Advanced Functional Capabilities*



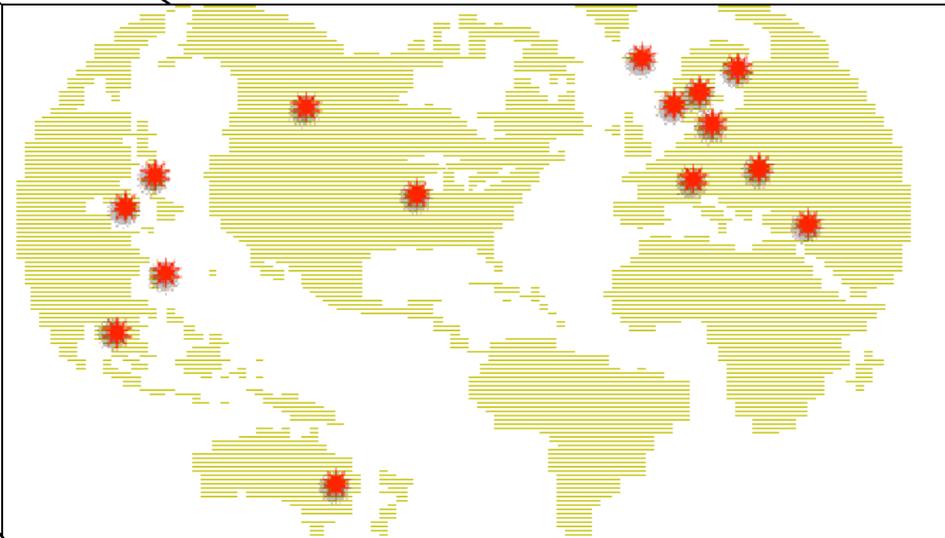
# Evolution of Understanding in a Distributed Knowledge Environment



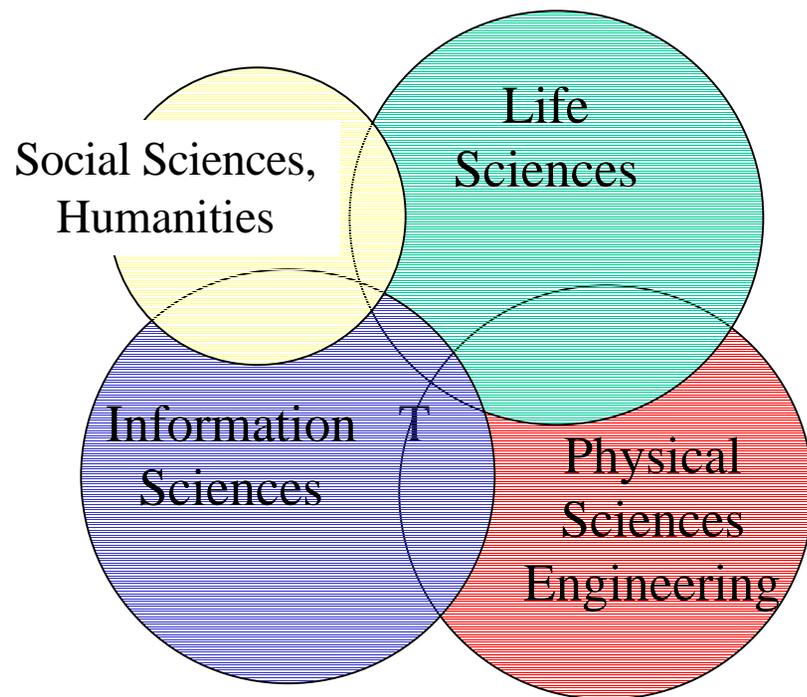
# *Changing Scales and Contexts of Interaction and Collaboration*



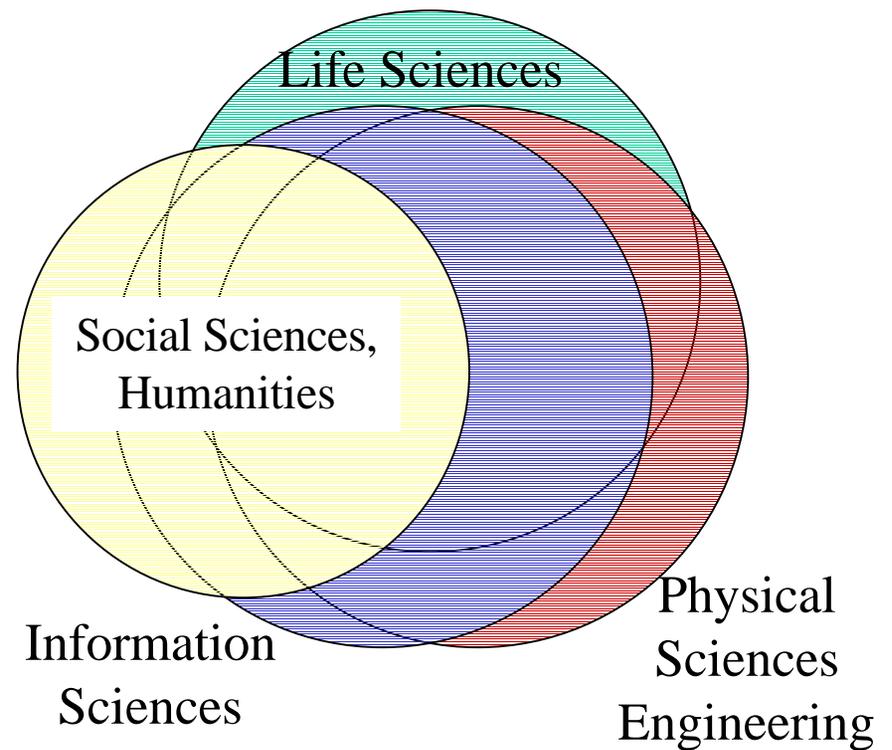
NSFNet StarTap Connections



# *A Vision of Disciplinarity: The World in 2010*

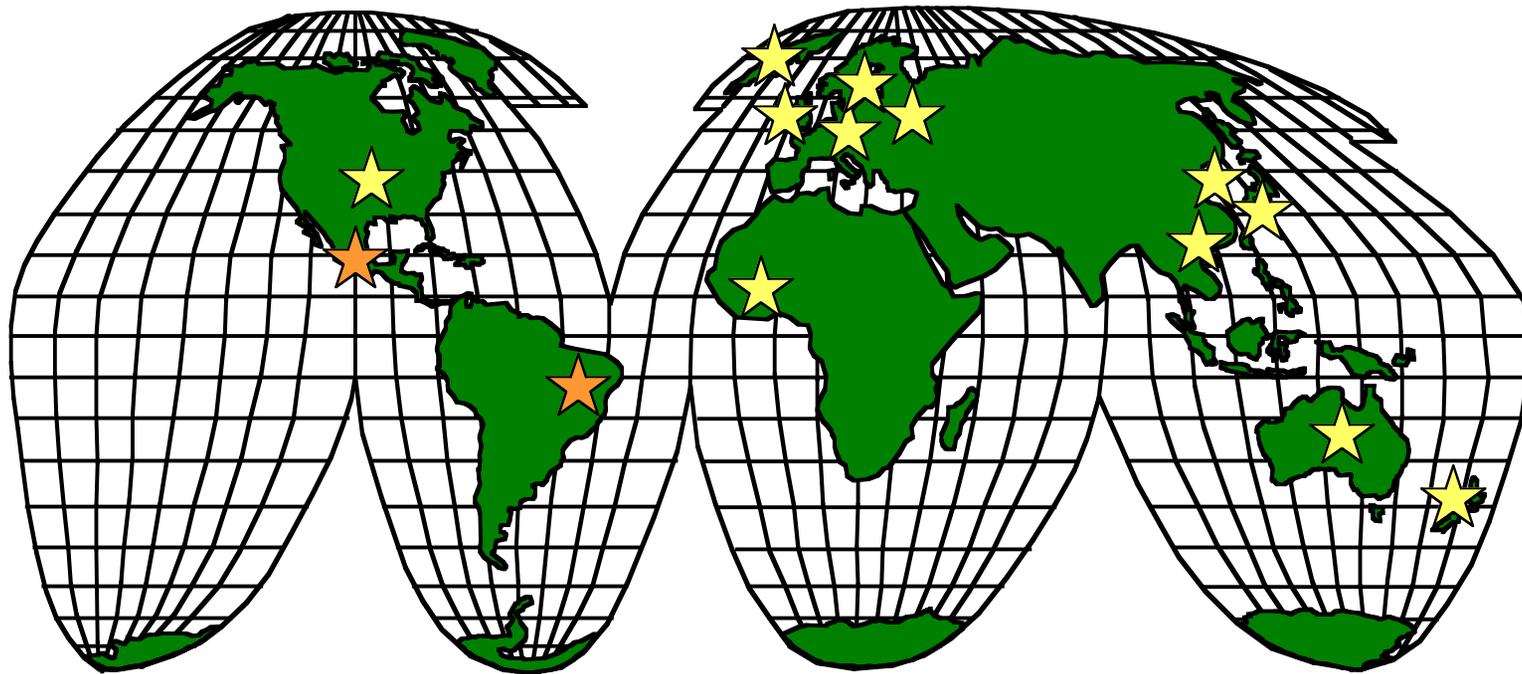


**2000**



**2010**

# *International Digital Libraries Collaborative Research Program*



## **FY 1999 Competition Data**

- ~50 proposals requesting \$25M
- ~30 countries
- **Formal Program with UK/JISC (Circular 15/98)**

# Languages and the Internet

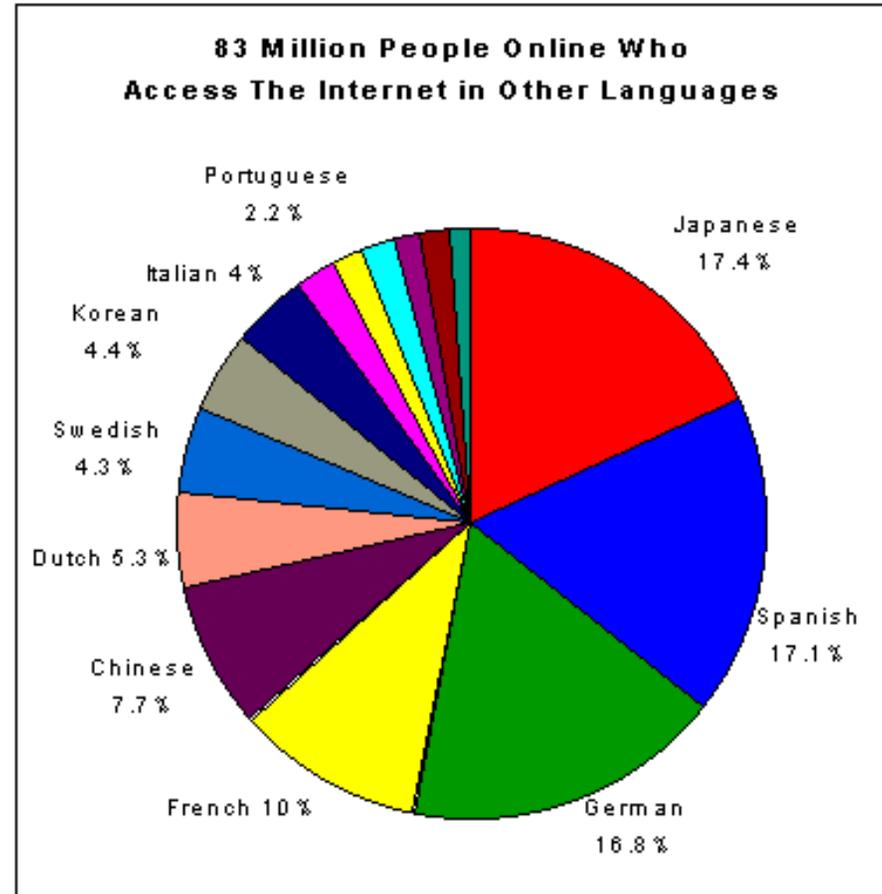


## April 1999

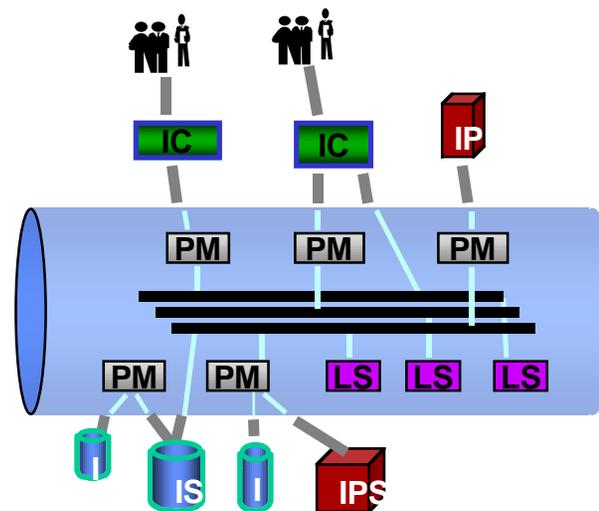
<b>English</b>	<b>107.2M</b>	<b>56.5%</b>
<b>non-English</b>	<b>82.7M</b>	<b>43.5%</b>
<b>European</b>	<b>54.9M</b>	<b>30.0%</b>

## By end of year 2000

<b>English</b>	<b>160M</b>
<b>non-English</b>	<b>167M</b>



# *Making Digital Libraries Infrastructure Requires Dealing with heterogeneity at Many Levels\**

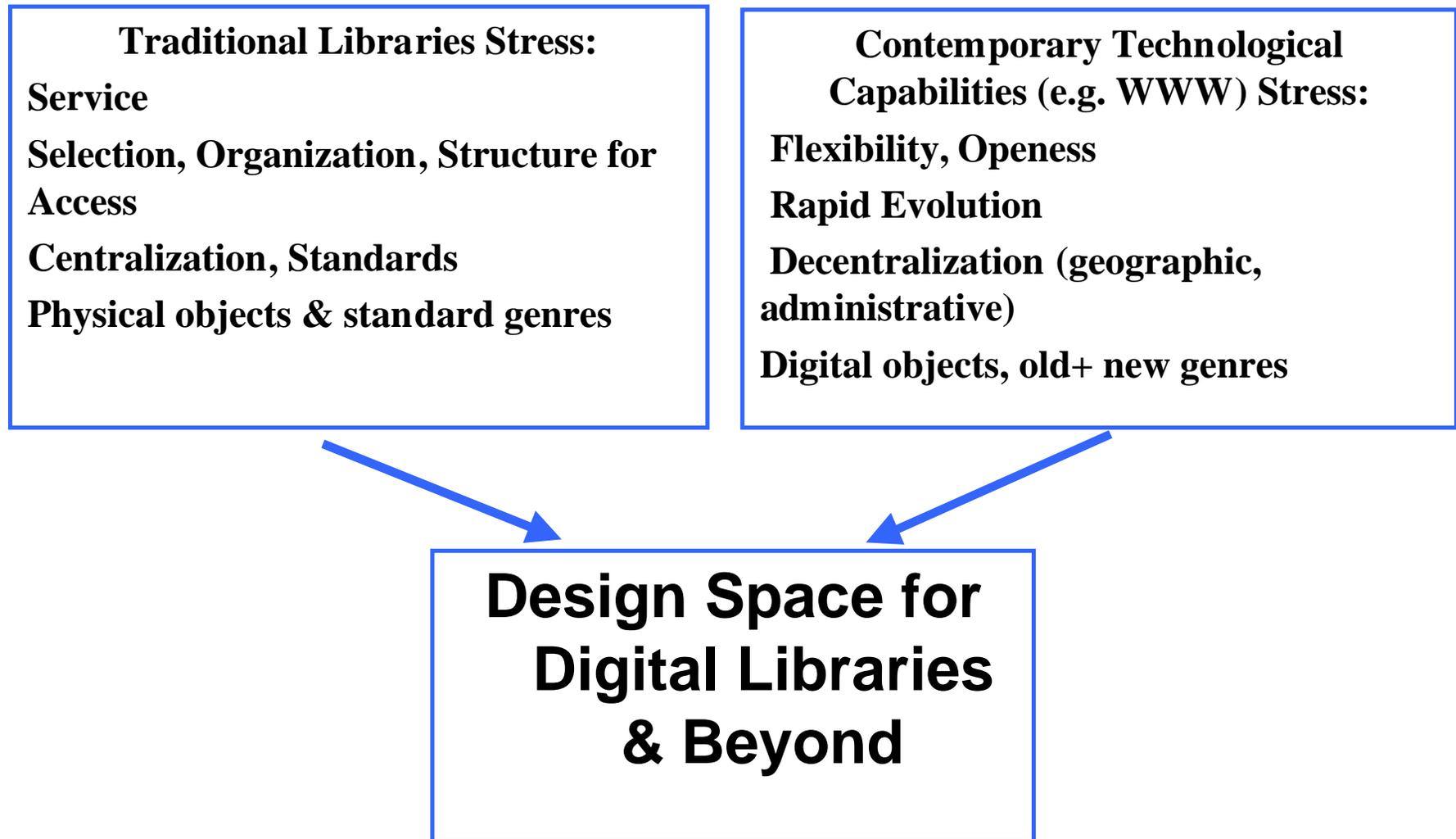


**PM:** Protocol Machine  
**LS:** Library Service  
**IC:** Interface Client  
**IS:** Information Source  
**IPS:** Information Processing Service

**Stanford InfoBus:** CORBA distributed object technology

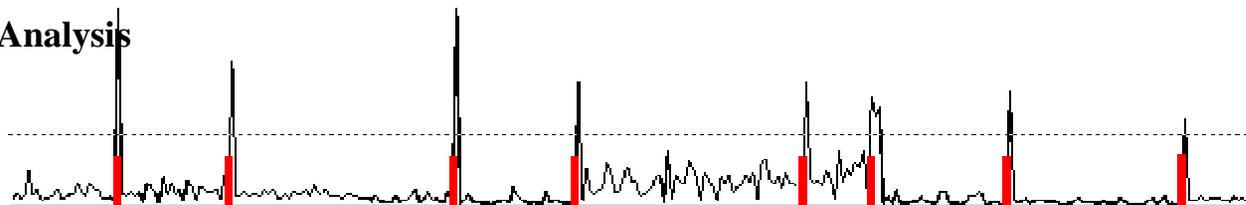
\* objects, collections, services, platforms....

# *Making Digital Libraries Infrastructure Requires Merging Intellectual Perspectives*



# *Making Digital Libraries Infrastructure Requires Application of Integrated Technologies*

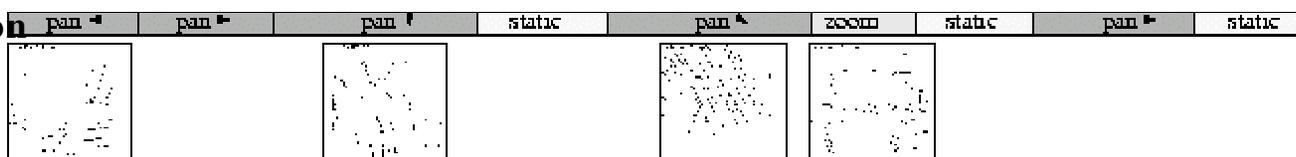
## Histogram Scene Analysis



## Scene Changes



## Camera Motion



## Word Relevance



## Key Words



## Audio Level



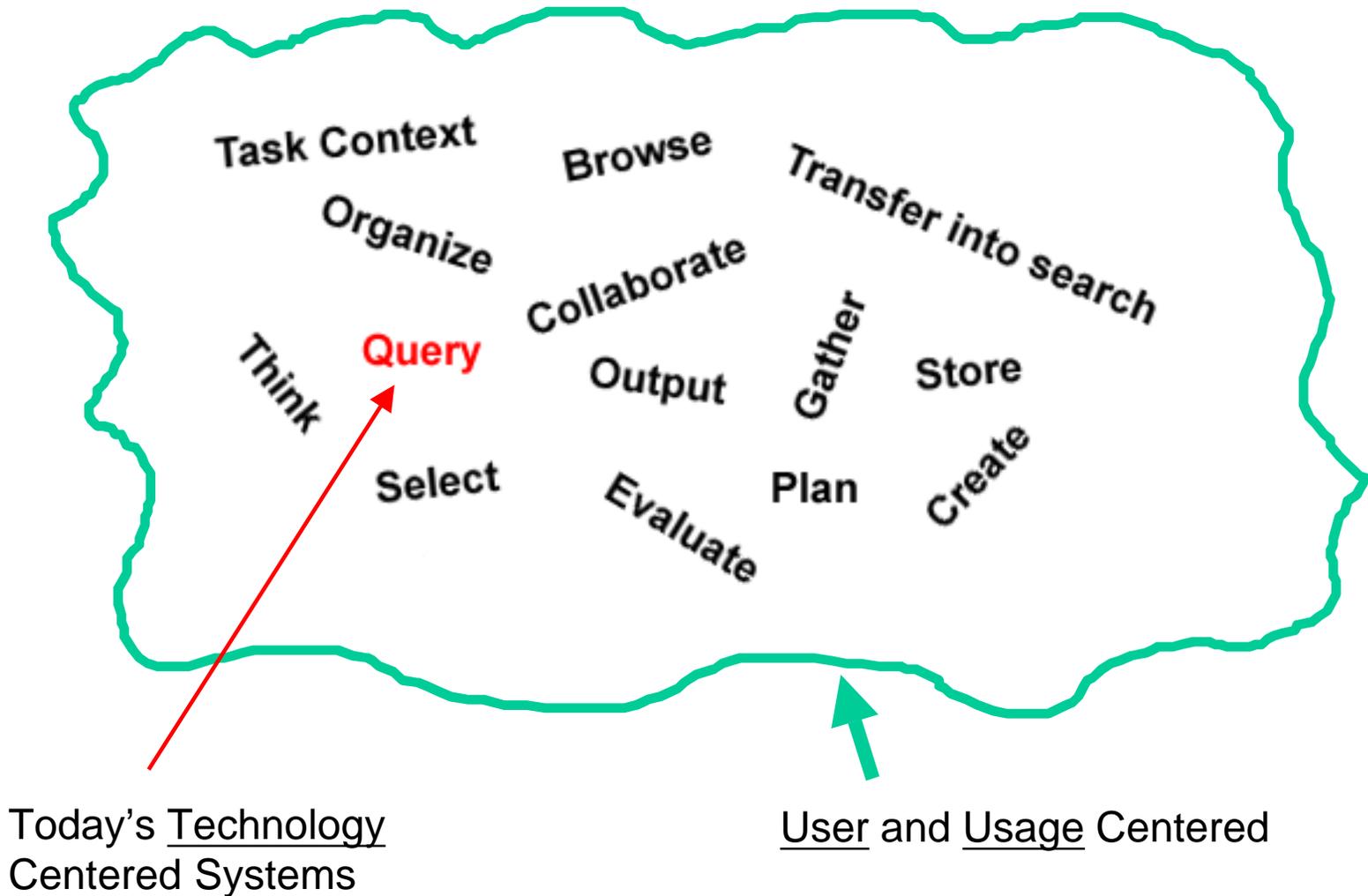
# *Making Digital Libraries Infrastructure Requires Building Large Collections of Diverse Information*



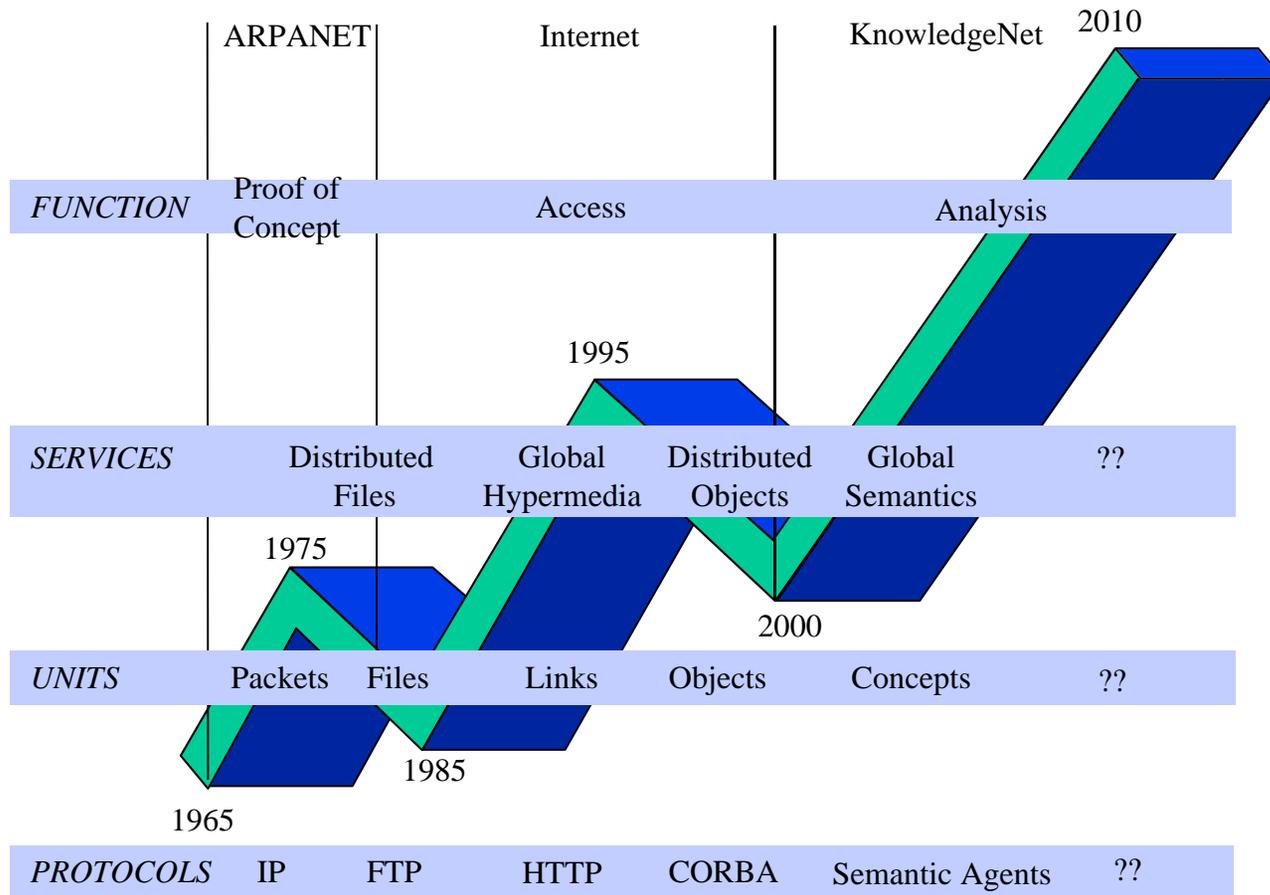
UC Berkeley Testbed

Type	Examples	June 96		Dec 96	
Documents	articles, EIRs, water reports	40,900 pp.	20 GB	96,600 pp	48GB
Images	DWR wildflowers Corel Habitats Total	14,838 2,905 22,000 0 39,743	238GB	15,506 7,437 28,101 158 52,000	306GB
Aerial photos	Suisun March Sac-SJ Delta	5000 img	3.4GB	500 img	3.4GB
Sensor Data	Delta fish flow	30 days	.02MB	30days	.02MB
GIS Data	dams, fish, watersheds, etc.	various	50MB	various	52MB
DOQs	SF Bay Area	102 img	5GB	102 Img	5GB
Digital Line Graphs	SF Bay North Coast				100MB 100MB
<b>Total</b>			<b>268GB</b>		<b>363GB</b>

# *Making Digital Libraries Infrastructure Means Supporting More than Query*



# Making Digital Libraries Infrastructure Requires New Conceptualizations of the Future (imagination)



# Goals for the Future

---



**Gather information and build collections**

**(to better use what we have and discover what is missing...)**

**Create new communities**

**(to communicate and collaborate)**

**Make technology disappear**

**(from our awareness and experience)**



## **For More Information:**

---

### **Digital Libraries Initiative National Homepage**

**<http://www.dli2.nsf.gov/>**

### **NSF-EU International Working Groups**

**[http://www.si.umich.edu/UMDL/EU\\_Grant/home.htm](http://www.si.umich.edu/UMDL/EU_Grant/home.htm)**

**<http://www.iei.pi.cnr.it/DELOS//NSF/nsf.htm>**

### **D-Lib Magazine**

**<http://www.dlib.org>**



---

*the definition of “digital library” continues to evolve...*

**Network accessible digital objects (representing text, documents, images, sound, video, agents, databases, middleware...) with sufficient identity, structure and contextual information to allow creating coherent collections on demand to service diverse user needs (query, analysis, communication, collaboration, ...)**

*S Griffin May 2000*

# DLI-2 Round 1 Awards



<b>DLI-2 Round 1 Awards</b>				
<b>Award ID</b>	<b>PI Name</b>	<b>Institution</b>	<b>Mos.</b>	<b>\$K</b>
9817485	Kornbluh, Mark	Mich St	60	3,600
9817484	Crane, Gregory	Tufts	60	2,758
9817434	McKeown, Kathleen	Columbia University	60	5,002
9817496	Wactlar, Howard D.	CMU	48	4,000
9817432	Smith, Terrence	UCSB	60	5,400
9817799	Garcia-Molina, Hector	Stanford University	60	4,300
9817353	Wilensky, Robert	UC-Berkeley	60	5,000
9874747	Verba, Sidney	Harvard University	36	1,800
9817416	Lagoze, Carl	Cornell University	48	2,268
9874759	Etzioni, Oren	U Washington	36	598
9817492	Gorman, Paul	Oregon Health Sciences	36	650
9817511	Weiderhold, Gio	Stanford University	36	520
9817430	Choudhury, Sayeed	Johns Hopkins	36	530
9874771	Armistead, Samuel G.	UC-Davis	36	497
9817483	Seales, W. Brent	U Kentucky	36	500
9817444	Buneman, Peter	U Pennsylvania	36	505
9874781	Rowe, Timothy	U Texas, Austin	36	500
9817527	Myers, Brad	CMU	36	450
9817473	Chen, HC	U Arizona	36	501
9817572	Palakal, M.	Indiana Univ	36	316
9817518	Willer, D.	U South Carolina	48	1,199
<b>Subtotal</b>				<b>40,894</b>

## DLI-2 Round 1 Awards - International



<b>DLI-2 International Awards (FY99)</b>				
<b>Award ID</b>	<b>PI Name</b>	<b>Institution</b>	<b>Mos.</b>	<b>\$K</b>
9975164	Larson, Ray	UC-Berkeley	36	305
9905842	Byrd, Donald	U Massachusetts	36	494
9905935	Hedstrom, Margaret	U Michigan	36	488
9906025	Calcari, Susan	U Wisconsin-Madison	36	480
9907892	Lagoze, Carl	Cornell U/ePrint	36	292
9905955	Lagoze, Carl	Cornell U/ILRT	36	240
<b>Subtotal</b>				<b>2,299</b>

# DLI-2 Round 1 Awards - Undergraduate Emphasis



<b>DLI-2 Undergraduate Emphasis Awards</b>				
<b>Award ID</b>	<b>PI Name</b>	<b>Institution</b>	<b>Mos.</b>	<b>\$K</b>
9817406	Agogino, Alice	UC-Berkeley	12	200
9816026	Maly, Kurt	Old Dominion Univ	12	80
9816644	Kappelman, John	U Texas, Austin	24	287
9816644	Druin, Alison	U Maryland	24	287
9980130	Owen, Scott	Georgia St	36	330
9980116	Agogino, Alice	UC-Berkeley	24	400
9979967	Wittenberg, Kate	Columbia U Press	36	581
9980049	Graves, William	Collegis Research	24	1143
<b>Subtotal</b>				

# DLI-2 Principal Investigator Departments



<b>DLI-2 Principal Investigator Departments</b>		
Anthropology	Biomedical Information	Classics
Computer Science	Economics	English
Fine Arts	Geography	Geological Sciences
Government	Electrical Engineering	Environmental Science
History	Information Management	Information Studies
Language Technology	Library & Information Science	Linguistics
Management Info. Systems	Medical Informatics	Political Science
Psychology	Religious Studies	Robotics
Sociology	Spanish	Teacher Education



## DLI-2 Technology Focus by Project

3-D Modeling	California Santa Barbara, Texas Austin	Linking	Cornell (intl – ePrint)
Access Control	California Berkeley	Log (Trace) Analysis	Oregon Health Sciences
Agents	Indiana Bloomington, Washington	Mobile Computing	Stanford
Archiving/ Preservation	South Carolina, U. Michigan (intl)	Multimedia Fusion	CMU , Columbia
Audio Retrieval	Johns Hopkins, Michigan State, U Mass Amherst (intl)	Natural Language Processing	Columbia
Classification, Clustering	Arizona	OCR	California Berkeley, John Hopkins
Data (Access) Services	Harvard	Parallel Processing	Arizona
Digital Video	CMU	Protocols	Stanford
Economic Models	California Berkeley, Stanford	Personalization	Columbia
Electronic Notebooks	California Berkeley	Provenance	Penn.
Federation	California Berkeley (intl), Cornell, U Wisconsin-Madison (intl)	Restoring Manuscripts	Kentucky
Geographic Info. Systems	California Santa Barbara	Speech Processing	California Davis, Michigan State
Images	California Berkeley, California Santa Barbara, Kentucky, Stanford, Texas Austin	Summarization	CMU , Columbia
Information Filtering	Indiana, Stanford	Text Analysis	Tufts
Information Visualization	CMU	Video Editing	CMU
Learning Contexts	California Santa Barbara Barbara		

# DLI-2 Content Focus by Project



Bibliographic Records	Arizona
Engineering Education	California Berkeley
EPrints	Cornell (intl ePrint)
Folk Literature	California Davis
Geo-referenced Info.	California Santa Barbara
Health Care	Oregon Health Sciences
Humanities	Tufts, Kentucky
Library Reference	Washington
Medical Images	Stanford
Mixtures of Media	California Berkeley (ntl), Cornell (intl ILRT)
Patient Records	Columbia
Sheet Music	Johns Hopkins, U Mass. Amherst (intl)
Skeletons	Texas Austin
Simulations	South Carolina
Social Science Data	Harvard
Speech	Michigan State
Video	Carnegie Mellon
Web	Arizona, Pennsylvania, Washington
X-ray CT Scans	Texas Austin