

Task #s Digitization

---

PHOTOCOPY  
PRESERVATION

## **Background for Meeting on Digitizing Federal Resources**

November 14, 1994

**The issue:** The federal government holds vast, unique collections critical for many educational and entertainment products. This has always been true, but the new computer and communication technology provides unprecedented opportunities for making enormous amounts of information available throughout the country in compelling, and relatively inexpensive formats. Enormous private investments are being made to exploit these opportunity as the nation's entertainment and educational system shifts to multi-media.

In spite of the large amounts of private funding available, and in spite of the fact that the administration has made dissemination of federal information a central priority, the rate at which federal holdings are being made available in digital form is painfully slow. Federal funding is inadequate and concern about intellectual property rights and other issues have blocked significant private investment.

Our goal is to break this logjam. The mission is to make as much of the information in federal holdings available to as many Americans as possible as rapidly as possible. Beyond this, virtually all options remain open.

**Defining terms:** The federal material that should be considered for public dissemination is now in a variety of forms: scientific and technical data, collections of physical objects, works of art, motion pictures, videos, audio recordings (everything from collections of bird songs to interviews with poets and recordings of early Jazz), photographs, texts, and manuscripts. Once captured in digital form, representations of this material can be copied, stored, and communicated at low marginal cost. The nature of the digital representations can vary and standards are evolving. Most modern technical data is already captured in digital form and most modern texts are prepared using digital word processing. Photographs can be captured in a variety of resolutions but detail fine enough to satisfy specialists can require enormous amounts of data per object. Advanced data compression schemes can be helpful. Standards for digital compression of movies and video material are advancing rapidly to satisfy commercial markets for high definition television. The least explored area is digitizing three-dimensional physical objects. Technology developed for engineering models and medical purposes makes it possible to capture precise dimensional information on an object, detailed photographs of the objects surfaces (including interior details where relevant). The engineering representation and photographs can be combined in a three-dimensional representation that can be rotated, manipulated, and otherwise examined conveying much of the information that could be provided by handling the physical object. This kind of data would make it possible, for example, to pick it a simulated version of Morse's telegraph, examine it from all sides, operate and even disassemble it.

In general, technical information in areas like astronomy, weather, and geology are made available by NASA, NOAA, and other technical agencies because of the research needs of scholars who have been taking advantage of advanced computers and communication for some years. Cultural and art collections, however, are poorly represented. The Smithsonian, the

National Park Service, the National Archives, and the Library of Congress are the principal holders of this material.

**Options:** Given the number of private firms that could benefit from easier access to federal collections, and the advanced state of the art in digital capture and dissemination of information, the primary barrier to rapid progress is institutional, not technical. In approaching this, it is important to agree that while an enormous digital collection of federal holdings is a critical resource, the real value of the collection must come by the new generation of publishers who will provide creative and useful ways to explore the material. The resource acquires meaning only when part of a creative educational or entertainment experience that gives it context and meaning. This "value added" activity should be provided almost entirely from private firms, not from public actions.

Options include:

1. ***Federal Funding:*** The Smithsonian and other federal agencies have had limited funds to digitize photographs and other materials and make them freely available on internet and commercial data bases like Compuserve for several years. Increased federal funding for this activity would ensure free, universal availability. It is unlikely that it could be funded. The Smithsonian alone has more than 120 million objects in its collections, 20 million photographs, and vast collections of movies and sound recordings. The Park Service has a collection about 1/3 this size. Even making the generous assumption that digitizing each object would average \$10 dollars, and that Congress would appropriate \$10 million per year, digitizing the Smithsonian alone would require more than 140 years. Moreover, an attempt by the federal government to proceed by itself would provide no incentive for producers to disseminate the information captured and would risk falling behind the state-of-the-art of commercial technology, providing information in formats uninteresting to private distributors.
2. ***Matching Funds From Charitable Institutions:*** The library of Congress has been able to increase federal funding by a factor of 4-5 by using federal funds to attract charitable grants. Significant funding may be available from Annenberg and other organizations interested in educational uses of new media. But charitable funding can also suffer the limits described for federal funding.
3. ***Establish a New Corporate Entity:*** A new for-profit firm, comparable to University Microfilm or C-SPAN could be started for the sole purpose of creating and disseminating federal digital holdings. The firm could be capitalized by some combination of corporate investments and charitable contributions. The holdings would not be freely available but available at a rate schedule which made the material in raw form available at comparatively low cost to individual students or libraries but more costly for firms intent on incorporating the information in value-added products. Priorities for digitization could be set by a board which in part reflected the interests

of commercial resellers. This scheme would have the advantage of low federal cost but would not meet the test of providing federal material at zero cost.

4. ***Allow Private Firms Digital Rights to Public Information:*** Firms could be provided temporary (eg. 5-year) rights to public information in return for funding digitization of the material. This scheme would also permit rapid digitization of holdings with little public cost but would place priorities entirely in the hands of private firms. (Federal funding could, of course, supplement the private investment). The scheme would also greatly limit options for public access to the information. Model contracts could alleviate some of the most obvious problems. To prevent monopoly control, contracts could require the firms to sell the digital information collected from public sources for a pre-arranged fee. Cream skimming could be avoided by requiring firms to digitize entire collections, not just the most commercially valuable items.
5. ***Others (including combinations of the above)***

To: Henry Kelly, Ed Fitzsimmons, Tom Kalil  
From: David Lytel  
Date: 16 July 1994  
Re: First shot at National Digital Library Initiative

If success really does have a thousand parents, then this idea is well on its way. At least five activities could claim paternity:

1. The Cabinet-level Education, Training, and Re-employment initiative wishes to include digitization of essential educational materials for network-based distribution.
2. The President proposed, and the heads of the other G7 nations accepted, that a ministerial-level group be convened in the fall to work on the global information infrastructure. Among the three ideas accepted for the agenda is a Global Digital Library initiative. Ron Brown will represent the US at this meeting.
3. Secretary of Education Riley has begun a task force within his department on the reauthorization of the Library Services and Construction Act, due to expire at the end of FY1994. Some key opinion leaders in the university and library communities have proposed we revise the Act to eliminate the construction title, which has been zeroed out in this and previous administrations for its minimal impact, only to be restored by Congress. A Library Services and Communications Act that reprogrammed this money (\$18M) for digitization would not be without controversy but is politically plausible because it is a shift in emphasis rather than a removal of resources. Every librarian in the country is suspicious of us if not outright angry that we have eliminated funding in FY95 for the titles of the Higher Education Act that support libraries.
4. Anita Jones wants to consider a Digital Heritage work group of the NSTC Committee on Information and Communications that would tie more closely the R&D work being conducted on digital libraries at NSF, ARPA and NASA to the needs of agencies such as the National Archives, the National Endowment on the Humanities, the Smithsonian Institution, the Library of Congress, the Government Printing Office, and the National Technical Information Service. This R&D activity is part of the HPCC program, which includes among its FY1994 milestones a start on "domain specific digital libraries, large scale databases/data sets, and user interfaces for focused research communities..." [HPCC FY95 Implementation Plan, p. 31]
5. The first volume of the IITF's Committee on Applications and Technology report entitled "Putting the Information Infrastructure to Work" identifies digital conversion and preservation as a significant unaddressed activity: "Due to the amount of material to be considered for digitization...and the number of problems associated with conversion (lack of image standards, selection and organization practices for digital materials, the expense of the process, the strategic problem of mutilating an item in order to digitize it easily, copyright issues), building an efficient model for the digitization of analog must be considered early in the transition. Similarly, material created and only existing in digital form is not being archived or preserved in an orderly fashion. The issues of collecting digital items with a view to long-term archiving and preservation, particularly those without broad market appeal, are of little interest to entities

interested in immediate economic reward. Archiving and preserving for posterity are largely being ignored at this time." [p. 95] The CAT is in the process now of gathering information from agencies on activities to be included in the IITF FY96 budget crosscut, which will identify agency funds in NII applications areas.

So how should we go about this?

1. Well, what specifically are we trying to accomplish?

We are asking for a clear relationship between agency expenditures and a number of materials (documents, images, audio, video, etc.) that can be made available to the public via the Internet, or the identification of funds that go to technology or standards development in the area of digitization so that we may present a good, better, best range of funding alternatives. Subject matter can range from the nation's cultural heritage to government documents and archives to specific training materials.

2. Where are related activities in the Federal budget that can be expanded?

I think the primary vehicle here is the Department of Education and the introduction of an Administration bill to reauthorize a new LSCA. The other significant expenditures include the IITA component of HPCC (primarily involving, as noted above, NSF, ARPA, and NASA) and current activities at NARA that archive by microfilming rather than by digitization. NEH activities such as awards to support standards activities (ie. the Text Encoding Initiative, the Committee on the Interchange of Museum Information, and the Art Information Task Force) could be expanded, as could grants for digitization (ie. Dead Sea Scrolls project, Denver Public Library Digitized Photographs Project, Making of America project, and Project Open Book).

3. What is the best mechanism for interagency coordination?

Pick one. I think the best model is that selected by the Electronics Working Group of the NSTC, which has as its "primary parent" the NSTC Civilian Industrial Technology Committee but has the Communications and Information Committee (CIC) as its "secondary parent." The new White House Learning and Technology Initiative can have the NSTC Committee on Education and Training as its primary parent and the CIC as its secondary parent, which will have visiting rights, pay half of the new initiative's college tuition, and give it away should the initiative ever get married.

4. Who do you need behind this?

Each of these agencies should be asked what they will have in their FY96 budget requests that can be part of the National Digital Library Initiative: Department of Education, The National Archives, the National Endowment on the Humanities, the Smithsonian Institution, the Library of Congress, the Government Printing Office, and the National Technical Information Service and the HPCC programs at ARPA, NSF, and NASA. White House participants must include Sally Katzen, head of the Office of Information and Regulatory Affairs or Bruce McConnell, head of the Office of Information Policy. Arati Prabhakar is working on the IITF crosscut and

has asked CAT members to help collect this information. It might be wise to find ways to support that activity, which will gather the requisite information. Or a meeting could be called, but however it is pursued it will require the input of the people below:

**Department of Education (ED)**  
400 Maryland Avenue, SW  
WDC 20202

Richard Riley, Secretary  
Madelein Kunin, Dep'y Secretary  
202-401-1000

Linda Roberts, Special Advisor  
Room 4015  
202-401-1444  
fax 401-3093  
lroberts@ed.gov

Marshall ("Mike") Smith  
Undersec'y and Policy Director  
202-401-3389  
fax 401-3085

Department of Education (ED)  
Office of Educational Research and Improvement (OERI)  
555 New Jersey Avenue, NW  
WDC 20208-5571

Sharon Robinson, Assistant Secretary

Laurence Peters, Sr. Policy Advisor  
Rm. 602c  
202-219-2050  
fax 219-1402

Arthur Sheekey, Sr. Program Analyst  
202-219-1930  
fax 219-1725

Ray Fry, Director of Library Programs  
202-219-2293  
fax 219-1725

**National Archives and Records Administration (NARA)**

7th and Pennsylvania Avenue, NW  
WDC 20408

Trudy Peterson, Acting Archivist

202-501-5130

fax 501-5244

Lead staff:

James Hastings 301-713-7096

Charles Dollar 501-5532

Ken Thibodeau, head

Center for Electronic Records

202-501-5575

kth@cu.nih.gov

**National Endowment for the Humanities (NEH)**

1100 Pennsylvania Avenue, NW

WDC 20506

202-606-8310

fax 606-8588

Sheldon Hackney, Chair

Lead staff:

Martha Chowning, Spec. Asst to Chairman

x-8355

George Farr, head of Div. of Preservation and Access

x-8570

**Smithsonian Institution**

WDC 20560

Robert Adams, Secretary of the Smithsonian

202-357-1846

Lead staff: James M. Hobbins, Exec. Assistant

Office of the Secretary

SI-215

202-357-1869

fax 786-2515



Marie Mattson, Director  
Office of Development  
A&I 1465  
202-357-2359  
fax 786-2516

David Allison  
Division of Computers, Information & Society  
202-633-9338  
fax 357-2038

**Library of Congress**  
WDC 20540

Dr. James Billington, Director  
202-707-5213

Lead staff:

Herb Becker, Director (Harriet, sec'y)  
Information Technology Services  
Madison Bldg, G-51  
202-707-6207  
fax 707-0955  
hbec@seq1.loc.gov

Jane Bortnick Griffith, Asst. Chief  
Science Policy Research Division  
CRS/SPR  
Mail code 7490  
Madison Building, Rm. 611  
202-707-9547  
fax 707-7000  
jbgr@seq1.loc.gov

**U.S. Government Printing Office (GPO)**  
732 North Capitol St., NW  
WDC 20401

Michael DeMario, Public Printer  
202-512-2034  
fax 512-1347

Lead staff:

Wayne Kelly, Superintendent of Documents (Asst Public Printer)  
202-512-0571

Judith Russell  
Director of Electronic Information Dissemination Service  
202-512-1622  
judy@eids01.eids.gpo.gov

**National Technical Information Service (NTIS)**  
5285 Port Royal Road  
Springfield, VA 22161

703-487-4600 main #

Don Johnson, Director  
703-487-4636  
fax 487-4093

Lead staff:

Ron Lawson, Associate Director and CFO  
703-487-4612

**Advanced Research Projects Agency (ARPA)**  
3701 North Fairfax Drive  
Arlington, VA 22203-1714

Gary Denman, Director  
703-696-2400  
fax 696-2209

Lead staff:

Duane Adams, Asst Dir.  
adams@arpa.mil  
703-696-2402

Howard Frank  
hfrank@arpa.mil  
703-696-2409

Steve Squires, Special Asst  
703-696-2262  
squires@darpa.mil

**National Aeronautics and Space Administration (NASA)**

Headquarters  
300 E Street SW  
WDC 20546

Dan Goldin, Administrator  
202-358-1010  
fax 358-2810

Lead Staff:

HPCC Office

Lee Holcomb  
Paul Hunter

202-358-4606 desk  
202-358-2747 sec'y  
fax 358-3557

**National Science Foundation**  
4201 Wilson Blvd  
Arlington, VA 22230

Neal Lane, Director  
703-306-1000

Lead Staff:

Not sure here, but Larry Rosenberg was head of digital library program but has passed away. He was in the CISE directorate's Division of Information, Robotics & Intelligent Systems, which is run by Y.T. Chen, 703-306-1930 or ytchen@nsf.gov

Lurthier Williams, Asst Dir for Education and Human Resources  
703-306-1600  
fax 306-0399

Nora Sabelli  
Program Director, Applications of Advanced Technologies  
703-306-1651 x-5888  
fax 306-0434

Henry Kelly, Ed Fitzsimmons, Tom Kalil  
om: David Lytel  
te: 16 July 1994  
First shot at National Digital Library Initiative

Success really does have a thousand parents, then this idea is well on its way. At least five initiatives could claim paternity:

The Cabinet-level Education, Training, and Re-employment initiative wishes to include digitization of essential educational materials for network-based distribution.

The President proposed, and the heads of the other G7 nations accepted, that a ministerial-level group be convened in the fall to work on the global information infrastructure. Among three ideas accepted for the agenda is a Global Digital Library initiative. Ron Brown will represent the US at this meeting.

Secretary of Education Riley has begun a task force within his department on the authorization of the Library Services and Construction Act, due to expire at the end of 1994. Some key opinion leaders in the university and library communities have proposed we revise the Act to eliminate the construction title, which has been zeroed out in this and previous administrations for its minimal impact, only to be restored by Congress. A Library Services and Communications Act that reprogrammed this money (\$18M) for digitization would not be without controversy but is politically plausible because it is a shift in emphasis rather than a removal of resources. Every librarian in the country is suspicious of us if not outright angry that we have eliminated funding in FY95 for the titles of the Higher Education Act that support libraries.

Anita Jones wants to consider a Digital Heritage work group of the NSTC Committee on Information and Communications that would tie more closely the R&D work being conducted at digital libraries at NSF, ARPA and NASA to the needs of agencies such as the National Archives, the National Endowment on the Humanities, the Smithsonian Institution, the Library of Congress, the Government Printing Office, and the National Technical Information Service. This R&D activity is part of the HPCC program, which includes among its FY1994 milestones to start on "domain specific digital libraries, large scale databases/data sets, and user interfaces for focused research communities..." [HPCC FY95 Implementation Plan, p. 31]

The first volume of the IITF's Committee on Applications and Technology report entitled "Putting the Information Infrastructure to Work" identifies digital conversion and preservation as a significant unaddressed activity: "Due to the amount of material to be considered for digitization...and the number of problems associated with conversion (lack of image standards, selection and organization practices for digital materials, the expense of the process, the strategic problem of mutilating an item in order to digitize it easily, copyright issues), building an efficient model for the digitization of analog must be considered early in the transition. Particularly, material created and only existing in digital form is not being archived or preserved in an orderly fashion. The issues of collecting digital items with a view to long-term archiving and preservation, particularly those without broad market appeal, are of little interest to entities

interested in immediate economic reward. Archiving and preserving for posterity are largely being ignored at this time." [p. 95] The CAT is in the process now of gathering information from agencies on activities to be included in the IITF FY96 budget crosscut, which will identify agency funds in NII applications areas.

So how should we go about this?

1. Well, what specifically are we trying to accomplish?

We are asking for a clear relationship between agency expenditures and a number of materials (documents, images, audio, video, etc.) that can be made available to the public via the Internet, or the identification of funds that go to technology or standards development in the area of digitization so that we may present a good, better, best range of funding alternatives. Subject matter can range from the nation's cultural heritage to government documents and archives to specific training materials.

2. Where are related activities in the Federal budget that can be expanded?

I think the primary vehicle here is the Department of Education and the introduction of an Administration bill to reauthorize a new LSCA. The other significant expenditures include the IITA component of HPCC (primarily involving, as noted above, NSF, ARPA, and NASA) and current activities at NARA that archive by microfilming rather than by digitization. NEH activities such as awards to support standards activities (ie. the Text Encoding Initiative, the Committee on the Interchange of Museum Information, and the Art Information Task Force) could be expanded, as could grants for digitization (ie. Dead Sea Scrolls project, Denver Public Library Digitized Photographs Project, Making of America project, and Project Open Book).

3. What is the best mechanism for interagency coordination?

Pick one. I think the best model is that selected by the Electronics Working Group of the NSTC, which has as its "primary parent" the NSTC Civilian Industrial Technology Committee but has the Communications and Information Committee (CIC) as its "secondary parent." The new White House Learning and Technology Initiative can have the NSTC Committee on Education and Training as its primary parent and the CIC as its secondary parent, which will have visiting rights, pay half of the new initiative's college tuition, and give it away should the initiative ever get married.

4. Who do you need behind this?

Each of these agencies should be asked what they will have in their FY96 budget requests that can be part of the National Digital Library Initiative: Department of Education, The National Archives, the National Endowment on the Humanities, the Smithsonian Institution, the Library of Congress, the Government Printing Office, and the National Technical Information Service and the HPCC programs at ARPA, NSF, and NASA. White House participants must include Sally Katzen, head of the Office of Information and Regulatory Affairs or Bruce McConnell, head of the Office of Information Policy. Arati Prabhakar is working on the IITF crosscut and

has asked CAT members to help collect this information. It might be wise to find ways to support that activity, which will gather the requisite information. Or a meeting could be called, but however it is pursued it will require the input of the people below:

**Department of Education (ED)**

400 Maryland Avenue, SW

WDC 20202

**Richard Riley, Secretary**

**Madelein Kunin, Dep'y Secretary**

202-401-1000

**Linda Roberts, Special Advisor**

Room 4015

202-401-1444

fax 401-3093

lroberts@ed.gov

**Marshall ("Mike") Smith**

**Undersec'y and Policy Director**

202-401-3389

fax 401-3085

**Department of Education (ED)**

**Office of Educational Research and Improvement (OERI)**

555 New Jersey Avenue, NW

WDC 20208-5571

**Sharon Robinson, Assistant Secretary**

**Laurence Peters, Sr. Policy Advisor**

Rm. 602c

202-219-2050

fax 219-1402

**Arthur Sheekey, Sr. Program Analyst**

202-219-1930

fax 219-1725

**Ray Fry, Director of Library Programs**

202-219-2293

fax 219-1725

**National Archives and Records Administration (NARA)**  
7th and Pennsylvania Avenue, NW  
WDC 20408

Trudy Peterson, Acting Archivist  
202-501-5130  
fax 501-5244

Lead staff:

James Hastings 301-713-7096  
Charles Dollar 501-5532

Ken Thibodeau, head  
Center for Electronic Records  
202-501-5575  
kth@cu.nih.gov

**National Endowment for the Humanities (NEH)**  
1100 Pennsylvania Avenue, NW  
WDC 20506

202-606-8310  
fax 606-8588

Sheldon Hackney, Chair

Lead staff:

Martha Chowning, Spec. Asst to Chairman  
x-8355

George Farr, head of Div. of Preservation and Access  
x-8570

**Smithsonian Institution**  
WDC 20560

Robert Adams, Secretary of the Smithsonian  
202-357-1846

Lead staff: James M. Hobbins, Exec. Assistant  
Office of the Secretary  
11-215  
02-357-1869  
ix 786-2515

Marie Mattson, Director  
Office of Development  
A&I 1465  
202-357-2359  
fax 786-2516

David Allison  
Division of Computers, Information & Society  
202-633-9338  
fax 357-2038

Library of Congress  
WDC 20540

Dr. James Billington, Director  
202-707-5213

Lead staff:

Herb Becker, Director (Harriet, sec'y)  
Information Technology Services  
Madison Bldg, G-51  
202-707-6207  
fax 707-0955  
hbcc@seq1.loc.gov

Jane Bortnick Griffith, Asst. Chief  
Science Policy Research Division  
CRS/SPR  
Mail code 7490  
Madison Building, Rm. 611  
202-707-9547  
fax 707-7000  
jbgr@seq1.loc.gov

U.S. Government Printing Office (GPO)  
732 North Capitol St., NW  
WDC 20401

Michael DeMario, Public Printer  
202-512-2034  
fax 512-1347



Lead staff:

Wayne Kelly, Superintendent of Documents (Asst Public Printer)  
202-512-0571

Judith Russell  
Director of Electronic Information Dissemination Service  
202-512-1622  
judy@eids01.eids.gpo.gov

**National Technical Information Service (NTIS)**  
5285 Port Royal Road  
Springfield, VA 22161

703-487-4600 main #

Don Johnson, Director  
703-487-4636  
fax 487-4093

Lead staff:

Ron Lawson, Associate Director and CFO  
703-487-4612

**Advanced Research Projects Agency (ARPA)**  
3701 North Fairfax Drive  
Arlington, VA 22203-1714

Gary Denman, Director  
703-696-2400  
fax 696-2209

Lead staff:

Duane Adams, Asst Dir.  
adams@arpa.mil  
703-696-2402

Howard Frank  
hfrank@arpa.mil  
703-696-2409

Steve Squires, Special Asst  
703-696-2262  
squires@darpa.mil

**National Aeronautics and Space Administration (NASA)**

Headquarters  
300 E Street SW  
WDC 20546

Dan Goldin, Administrator  
202-358-1010  
fax 358-2810

Lead Staff:

**HPCC Office**

Lee Holcomb  
Paul Hunter

202-358-4606 desk  
202-358-2747 sec'y  
fax 358-3557

**National Science Foundation**  
4201 Wilson Blvd  
Arlington, VA 22230

Neal Lane, Director  
703-306-1000

Lead Staff:

Not sure here, but Larry Rosenberg was head of digital library program but has passed away. He was in the CISE directorate's Division of Information, Robotics & Intelligent Systems, which is run by Y.T. Chen, 703-306-1930 or ytchen@nsf.gov

Lurthier Williams, Asst Dir for Education and Human Resources  
703-306-1600  
fax 306-0399

Nora Sabelli  
Program Director, Applications of Advanced Technologies  
703-306-1651 x-5888  
fax 306-0434