# IEEE Massive Storage Systems Sustainability of Digital Information

Elizabeth Cohen, Ph.D.
Vice Chair Education Academy of Motion Picture Arts and Sciences

Blue Ribbon Task Force on Sustainable Digital Preservation and Access

# Welcome



The Andrew W. Mellon Foundation

Council on Library and Information Resources

JISC

*We greatly appreciate the funding, time, and support from our sponsors.*

Cohen Acoustical

NARA

Rensselaer
why not change the world?

OCLC

University of Virginia

UCL

DCC

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

UCIrvine

cni
Coalition for Networked Information

Smithsonian Institution

Berkeley
University of California

Microsoft

Johns Hopkins University

Michigan

AK Consulting

# Why this Tutorial?

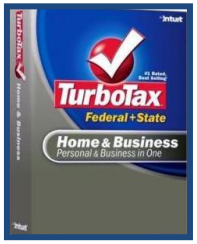
- **Move the conversation** from the Blue Ribbon Task Force to the broader community
- **Do something now** – get ahead of the problem of digital data loss and degradation
- **Focus on concrete actions** for key common scenarios:
  - Research Data
  - Scholarly Discourse
  - Collectively-produced web content
  - Commercially-owned cultural content

# We Depend on Digital Information

DATA.GOV

E-Government

Microsoft




Google

E-Business

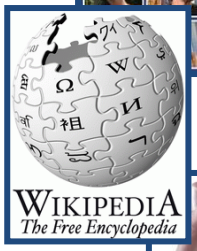




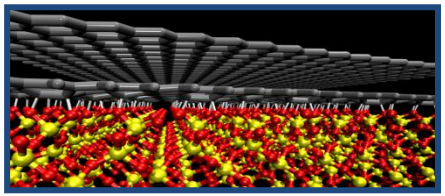


Digital Entertainment

Research and Education




Communication and Information

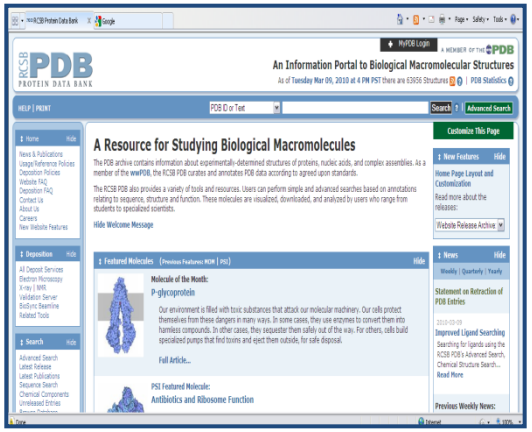


# Access to Information Tomorrow Requires Preservation *Today*

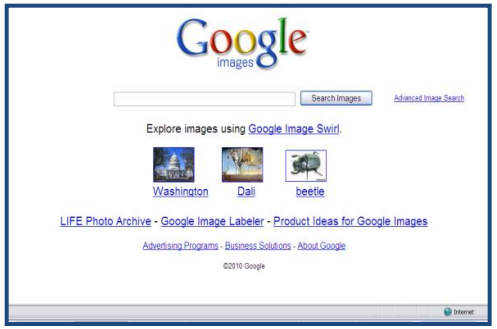
- **Digital Access and Preservation** is a technical, management, policy, regulatory, social, and **economic** problem

- Key issues to resolve:
  - **What should we preserve?**
  - **Who is responsible** for digital information?
  - **Who pays** for digital information and its supporting cyberinfrastructure?

# How do we currently support access to digital information?



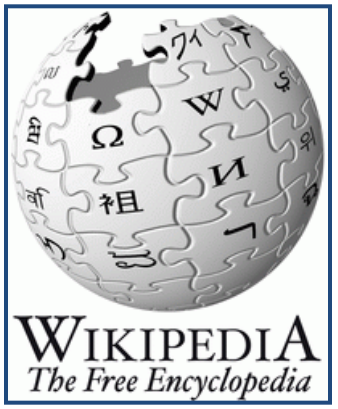

Federal grants




Advertisements


Subscription

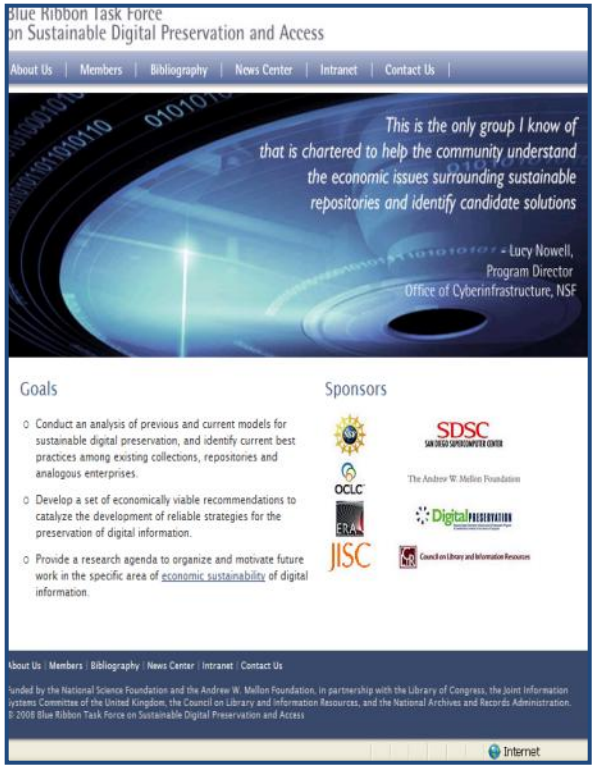

Pay per service




Donations, etc.

# Focus on the Long-term: The Blue Ribbon Task Force on Sustainable Digital Preservation and Access

- **BRTF Charge:**
  1. Conduct a comprehensive **analysis** of sustainable digital preservation
  2. Identify and evaluate **best practices**
  3. Make specific **recommendations for action**
  4. Articulate **next steps** for further work




brtf.sdsc.edu

# Who's Involved: BRTF-SDPA Participants

## Blue Ribbon Task Force:

- Fran Berman, Rensselaer Polytechnic Institute [co-Chair]
- Brian Lavoie, OCLC [co-Chair]
- Paul Ayris, University College London
- Sayeed Choudhury, Johns Hopkins University
- Elizabeth Cohen, AMPAS/Stanford
- Paul Courant, University of Michigan
- Lee Dirks, Microsoft
- Amy Friedlander, CLIR
- Vijay Gurbaxani, UC Irvine
- Anita Jones, University of Virginia
- Ann Kerr, Consultant
- Cliff Lynch, CNI
- Dan Rubinfeld, UC Berkeley
- Chris Rusbridge, DCC
- Roger Schonfeld, Ithaka
- Abby Smith Rumsey, Consultant
- Anne Van Camp, Smithsonian

## Liaisons

- NSF: Chris Greer, Lucy Nowell, Sylvia Spengler, Phil Bogden
- Mellon Foundation: Don Waters
- Library of Congress: Laura Campbell, Martha Anderson
- NARA: Robert Chadduck

## Staff and Students

- Task Force Administration: Susan Rathbun
- Communications: Jan Zverina, Ben Tolo
- Graduate Student Interns: Lori Eakin, Liz Bedford

# BRTF Deliverables

December 2008: **Interim Report**

- Foci:
  - Understand current practices
  - Synthesize major themes
  - Identify systemic challenges



February 2010: **Final Report**

- Foci:
  - Structural analysis of 4 common scenarios
  - Recommendations and Actions
  - Next steps

# The BRTF Interim Report: Exploring the Landscape

- What are the requirements for “economic sustainability” of digital information?
- What are the roadblocks?
- What are the current challenges?
- What are best practices?

# “Economic Sustainability” requires …

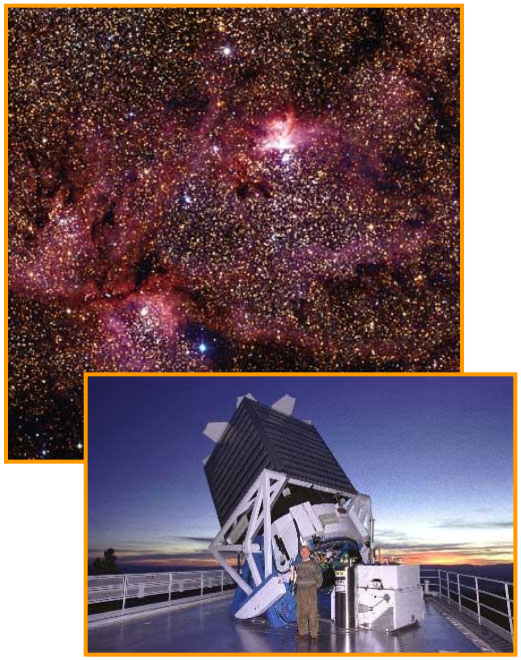
- **Recognition of the benefits** of long-term access and preservation
- **Incentives** for decision-makers to act
- **Means of selecting “valued” information** for long-term preservation

- Mechanisms to support **ongoing, efficient allocation of resources**
- Appropriate **organization and governance** of preservation and access activities

** From Blue Ribbon Task Force Interim Report*

# What are the roadblocks? Systemic Challenges to Digital Preservation

- **Long-term preservation funded by short-term allocations**
- Lack of alignment between stakeholders, roles, and responsibilities
- **Inadequate institutional, enterprise and/or community incentives**



*Findings from Blue Ribbon Task Force Interim Report*

# Alignment Challenges: The Stakeholder Problem

- **Many Stakeholders in digital preservation …**
  - **Stakeholders who benefit** from use of the preserved asset
  - **Stakeholders who select** what to preserve
  - **Stakeholders who own** the asset
  - **Stakeholders who preserve** the asset
  - **Stakeholders who pay**



- ***The greater the alignment between key stakeholder groups, the better the prospects for sustainable preservation***

*Findings from Blue Ribbon Task Force Final Report*

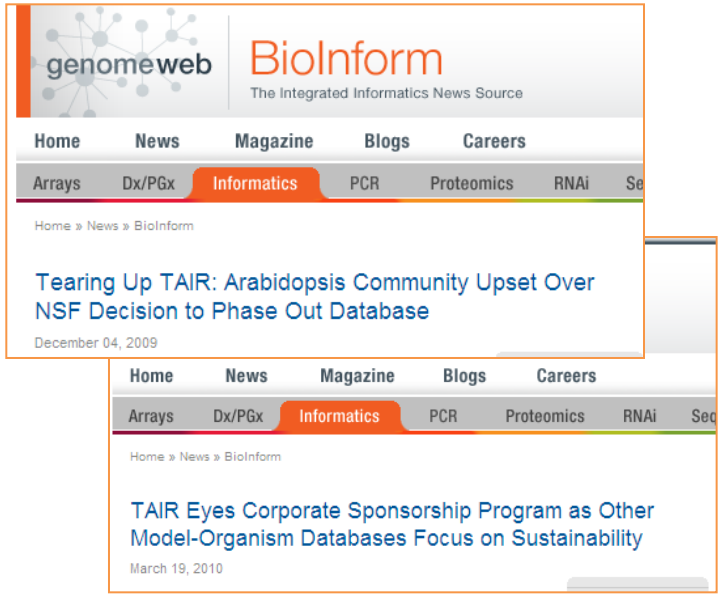
# Challenges for Stakeholders

- **Prioritizing the value of digital preservation** over more pressing short-term objectives
- **Complacency that current practices are good enough**
- **Fear that digital access and preservation is too big to take on.** Digital preservation is as big as the Information Age and as small as the resources you have control over…

*Findings from Blue Ribbon Task Force Interim Report*

# Funding and Incentives – Lessons Learned

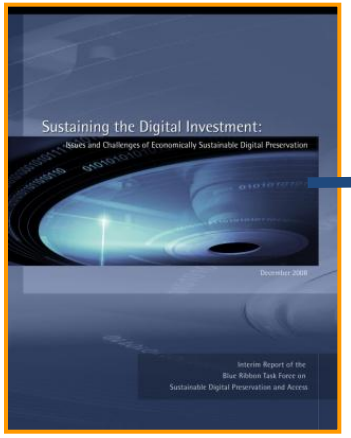
- **Diversity of funding streams is critical**
- **Non-monetary incentives are important.**
  - Corporate recognition and reputation enhancement are also incentives




- **Multiple options to consider when selecting an economic model to support digital preservation.**
  - *Think out of the box:* Public-spirited, mission-driven institutions sometimes resistant to monetizing and charging for a "social good"

*Findings from Blue Ribbon Task Force Interim Report*

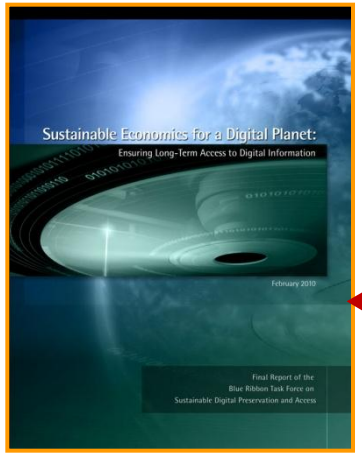
# Context to Findings




**Interim Report: Problem Space, Experience**

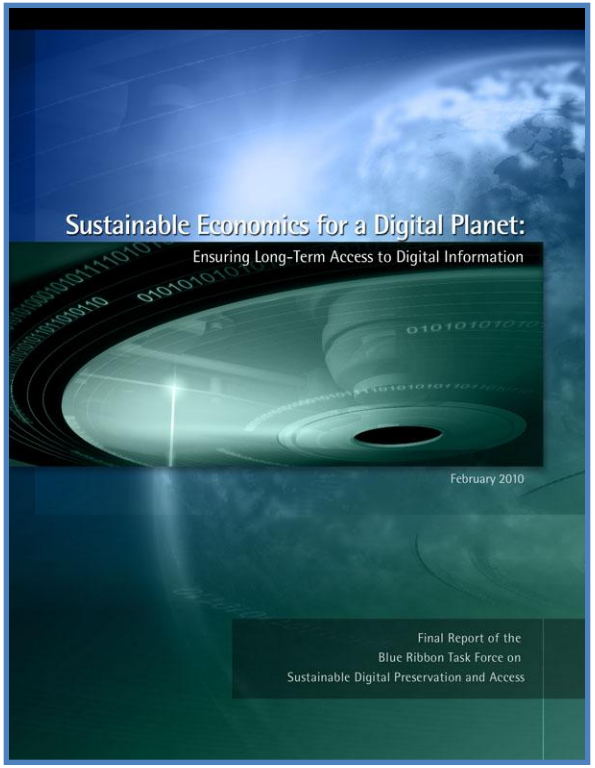
**Economic Framework 4 Common Scenarios**

**Final Report: General Findings & Recommendations**



# BRTF Final Report:

# *Sustainable Economics for a Digital Planet*



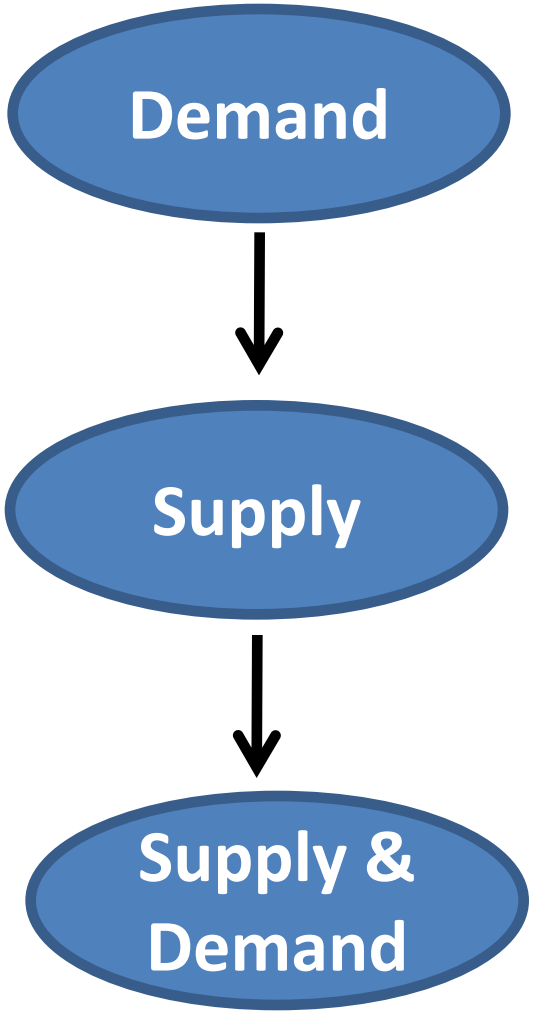

**http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf**

# Key Finding

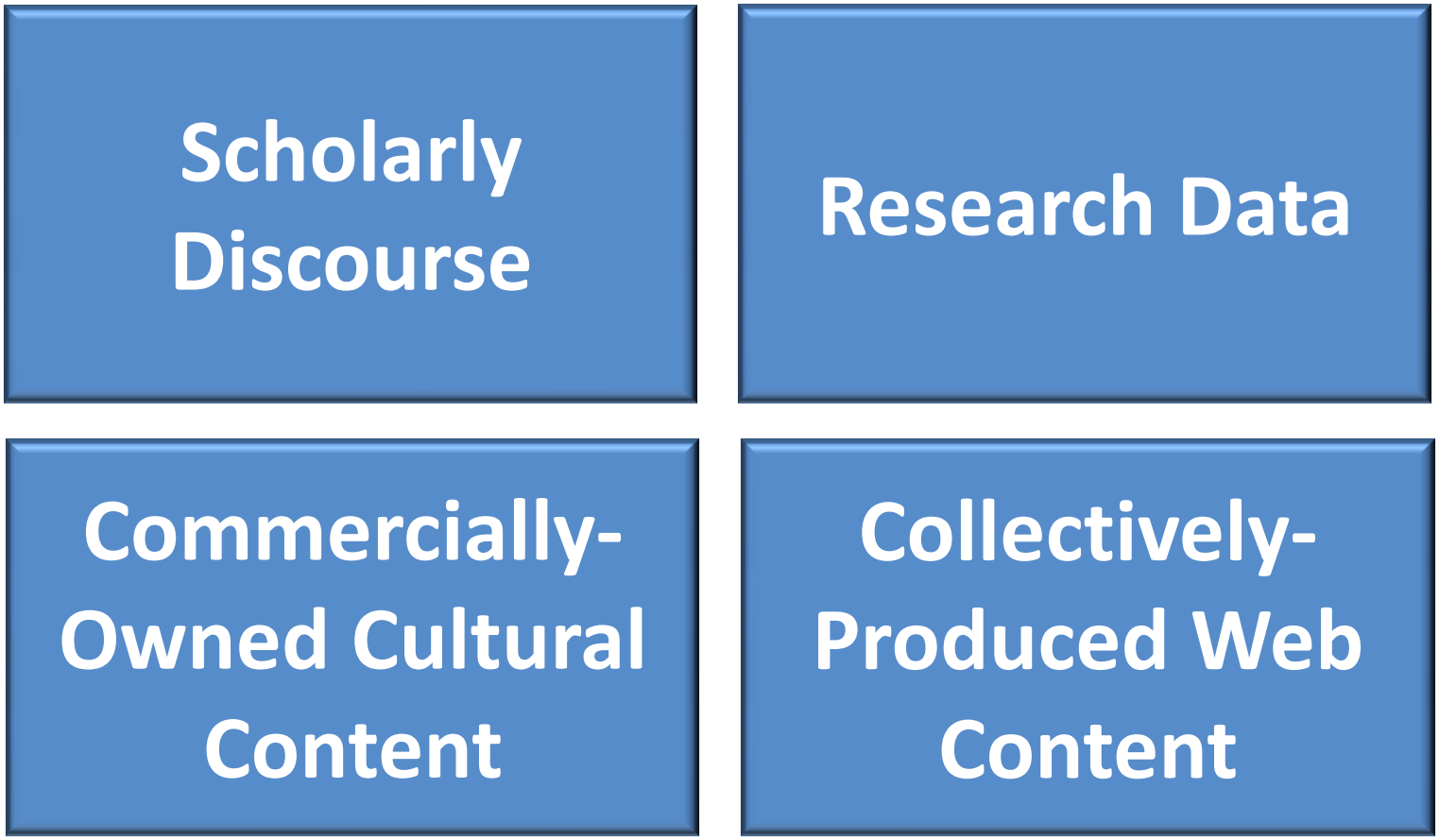
***"… sustainable economics for digital preservation is not just about finding more funds. It is about building an economic activity firmly rooted in a compelling value proposition, clear incentives to act, and well-defined preservation roles and responsibilities."***

# Key Themes

- **Value**
  - Expected benefits
  - Judgment about priorities
- **Incentives**
  - Motivation to act
  - Vary across stakeholders & time
- **Roles & Responsibilities**
  - Who benefits? Preserves? Pays?



# Digital Preservation Contexts



# Scholarly Discourse

## Sustainability Challenges

Consensus about value, (less so for emerging forms)

Incentives can be misaligned

Roles & responsibilities rooted in print era

## Actions

Selection criteria for emerging forms

Diffuse right to preserve to those willing to act in the public interest

Align responsibilities with digital environment

# Research Data

| Sustainability Challenges | Actions |
| --- | --- |
| Vast amounts of data; uncertain future value,<br><br>Incentives diminish as decision-making becomes more “granular”<br><br>Fragmented decision-making/preservation | Establish priorities in data selection<br><br>Leading role for funders in mandating preservation<br><br>Coordination and scale can leverage value and reduce costs |

# Commercially-Owned Cultural Content

## Sustainability Challenges

Variable/diffuse demand; "digital cultural heritage" uncertain

Private & public incentives to preserve often competing

Lack of "hand-off" mechanisms

## Actions

Define digital cultural heritage to aid selection

Establish public right to preserve that protects private interests; enhance private incentives

Public-private partnerships to ensure lifecycle curation

# AMPAS Grand Challenge

- A new approach to digital storage technology that is economically and technologically sustainable for 100 years or longer and survives periods of benign neglect.

# Valuable Information

- In 2008 Americans consumed 36 Zettabytes( 1 zetabyte equals a million million gigabytes) or 34 gigabytes of digital information everyday for every man, woman, and child in the US.
- More than 25% of  this “born digital” data is viewed by Americans as “priceless, impossible to replace”.

# Film

- Completely replaced by digital storage technologies for consumer photography, newsgathering operations, television production, and for an increasing portion of motion picture production and distribution chain.
- Today's digital storage technology is not designed with long term preservation as a fundamental requirement.

# Avatar

- Q: How is “sustainable digital preservation an issue for creatives, for example, how much data was created on Avatar, what different kinds of data, how was metadata handled?

# Long Term Storage

- Q: What infrastructure needs to be in place so that movies can be made “archives-ready”

# Titanic Disaster

- Q: What was the preservation plan for Avatar data?  What about Titanic?

# Major Challenges

- Q: What are the major economic challenges for audiovisual preservation in the next 5-10 years?

# Responsibilities

- Q: What are our preservation responsibilities as a content creator?

# Next Steps

- Q: How do we, as a society, collectively ensure that important cultural content that does not have commercial value is preserved for future users?

# Policies

- Q: Are there public policy changes or incentives that could facilitate those partnerships?

# Collaboration Invitation

- Q: As a content creator, do you have any suggestions on how your community can help with this?

# Collectively-Produced Web Content

## Sustainability Challenges

Future demand not clearly articulated

Incentives to preserve are weak

Ownership diffused; no clear preservation actor among current stakeholders

## Actions

Gather stakeholders under leadership of stewardship organizations

Use public policy to strengthen incentives

Contributors and hosting services should lower barriers to 3rd-party archives

# Some General Principles

- **Dynamics:** **Preservation is a *series* of decisions**
- **Benefits:** **Value of preservation based on use**
- **Selection:** **Scarce resources = prioritization**
- **Incentives:** **Strengthen, align, create**
- **Organization/Governance:** **Responsibilities must be clear; “handoffs” secure**
- **Resources:** **reflect community norms, flexible in face of disruptions, leverage economies of scale & scope**

# Concluding Thoughts

- **Different levels of progress across scenarios in thinking about sustainability issues**
- **Choices & trade-offs key to preservation decision-making:**
  - **e.g., creation vs. preservation**
  - **e.g., open benefits vs. limited benefits**
- **Scope for action for all types of stakeholders (institutions, funders, policy, individuals)**

# Initiatives

- **Initiative 1: Create deposit provisions for publications and other scholarly records that arise from federally funded research.**
- **Initiative 2: Create a national strategy that ensures that the knowledge supporting research and innovation today is available for the future.**
- ***The creation of demand:*** Current efforts from OSTP and federal funding agencies to mandate data management and stewardship plans for federally funded R&D are creating a need for economically sustainable archival services and institutions. The Task Force encourages OSTP to continue on this trajectory, which will result in the preservation of valuable digital information on which future innovation can be based.
- ***The creation of supply:*** There is a current gap in the capacity, cyberinfrastructure, and operational workforce needed to host the increasing of digital research data. The Task Force advocates the development of a program at the $100M or greater level to create "ramp-in" funding for institutions willing to provide 3rd party archival services for research data.
- Funding in the program would support necessary cyberinfrastructure capital investment and 5 years of operation, after which the awardees would be required to have an economically sustainable plan for an additional 10 years of archival services when ramp-in funding ends. Funding for this program should support archival services from both the public sector (including university libraries and 3rd party institutional repositories) and the private sector. Funding should also be widely distributed so that each state can begin to build archival capacity and services for its federally funded research and researchers.

# Initiatives

- **Initiative 3: Provide preservation funds for public-private partnerships to create sustainable archival storage solutions.**
- Provide preservation funds in the form of "innovation budgets". These would be used for public-private partnerships to develop research roadmaps for next generation archival storage systems and support a range of sectors ranging from government to higher education to the arts. This funding should emphasize translational research from higher education and "horizontal" cross-cutting technology platforms that may be beyond the scope of any individual firm. Partners should include established organizations, but also newer entrants into the digital arena. Within 12 months, these partners would be expected to develop well-defined business plans for ongoing operations that outline potential markets and impacts on national priorities such as IT innovation, green computing and workforce development. This recommendation reflects needs identified through projects involved in the NDIIPP and DataNet programs. OSTP could begin this process by convening a leadership group from higher education, storage vendors, and the entertainment industry to define the specific elements of the program and the goals for next-generation archival storage.
-

# Initiatives

- **Initiative 4: Develop greater cross-agency coordination for the purposes of digital access and preservation.**
- Because of the scale and complexity of the digital preservation challenge, there is a continuing need to coordinate developments and share information at many different levels. Within the Federal Government, agencies that fund external research need to coordinate policies and practices; all executive-branch agencies need to be concerned with the ongoing preservation of their own materials, in partnership with the National Archives and Records Administration (NARA), and the Government Printing Office (GPO). Legislative and Judicial branch organizations also face massive preservation challenges, and organizations like the Library of Congress, the National Library of Medicine and the National Library of Agriculture have important roles to play. The areas of concern range from research and technical deployment to economic, legal and public policy; there's an ongoing need to share knowledge, and to coordinate investments and activities.
- Looking beyond activities internal to the federal government, there's a strong need for outreach and assistance aimed at state and local government, as well as the need to coordinate federal activities with work going on in cultural memory organizations (libraries, archives, museums, local history societies, public broadcasting and the like), and in the higher education and research communities.