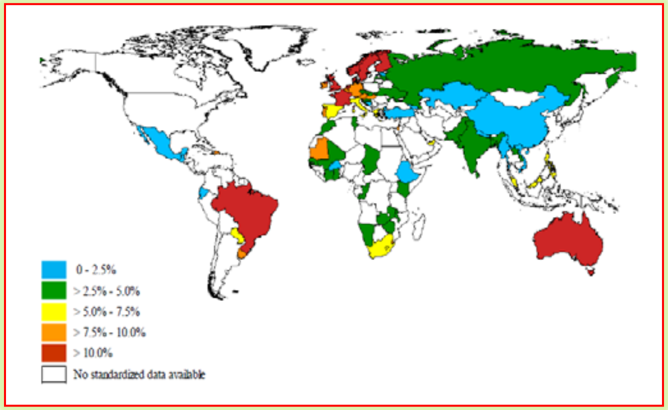
Building an Ecosystem to Accelerate Data-Driven Innovation

Dr. Francine Berman

Chair, Research Data Alliance / US

Edward P. Hamilton Distinguished Professor of Computer Science, Rensselaer Polytechnic Institute

Research Data Driving Solutions to Complex Scientific and Societal Challenges



Who is most at risk to contract asthma?

How can we increase wheat yields?

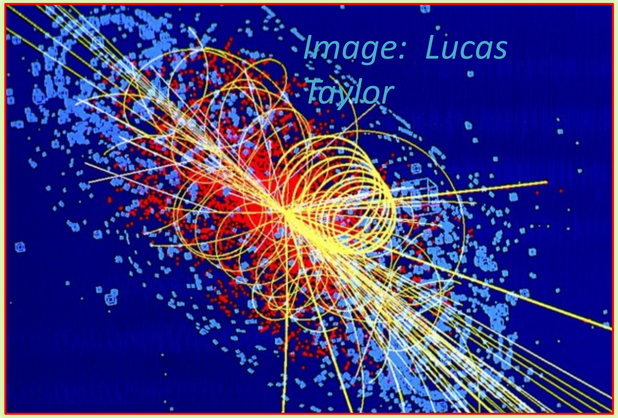


Image: Lucas Taylor

How accurate is the Standard Model of Physics?

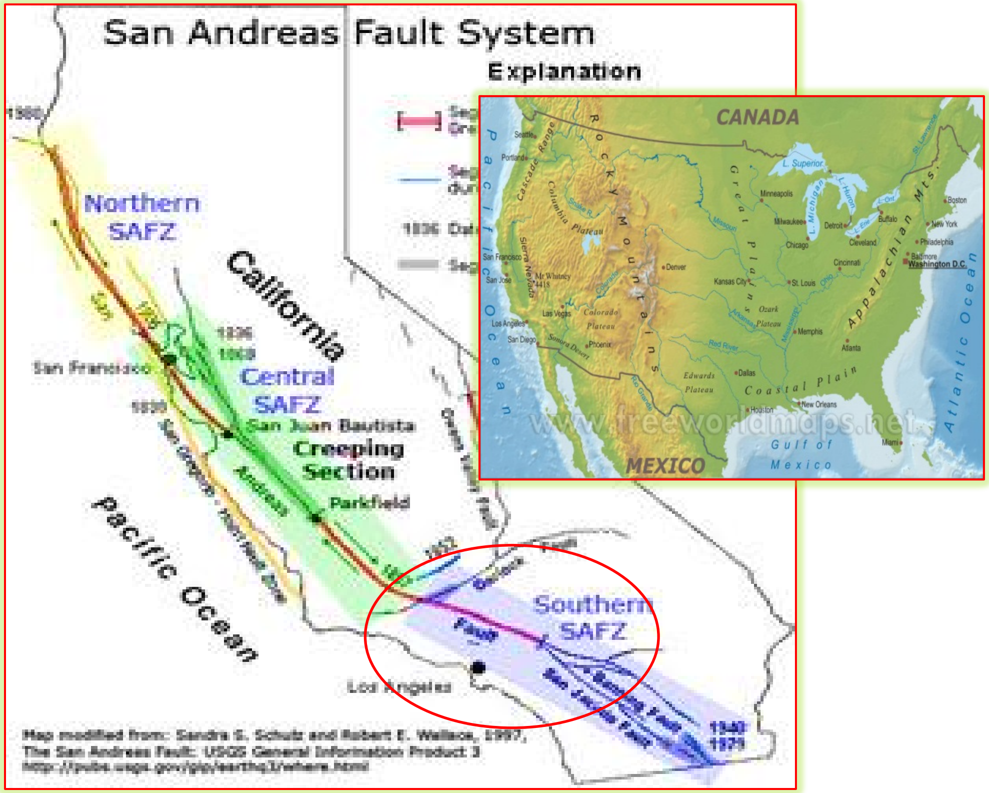
Image: Ceinturion, Wikipedia

How can we best address energy needs and sustain the environment?

Data-Driven Geoscience: How Can We Respond to Large-Scale Earthquakes?

Earthquake simulations enable

- Enhanced **scientific understanding** of the physical world
- More strategic plans for bridge, building and other physical infrastructure reinforcements to **increase safety**
- Better **disaster response planning** for police, fire fighters, ER teams in high-risk areas to increase their effectiveness



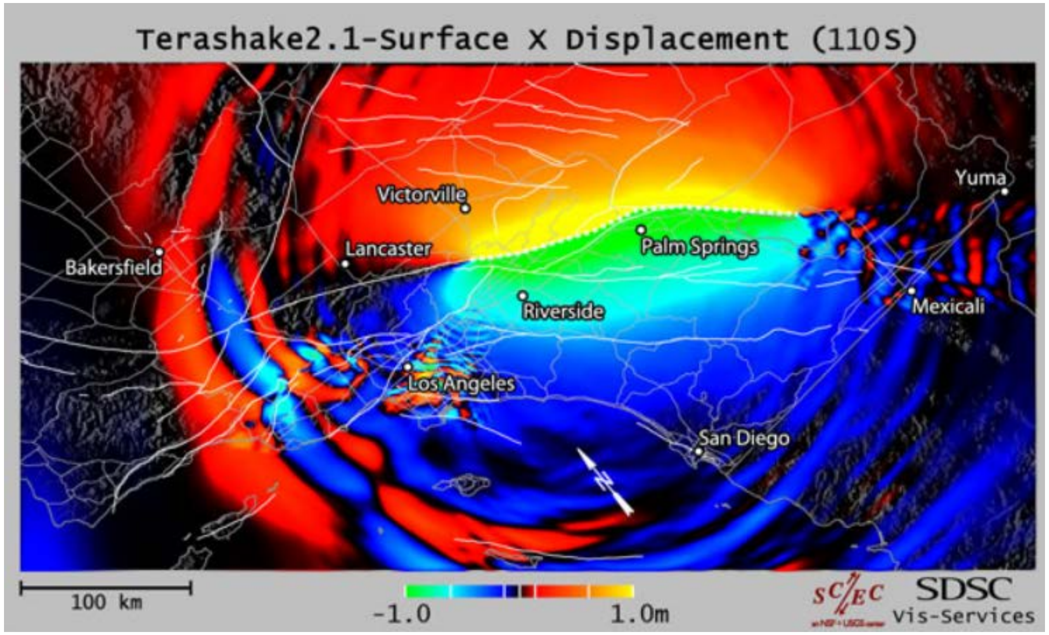
Simulation courtesy of Amit Chourasia, SDSC, Table information courtesy of Southern California Earthquake Center

Fran Berman

TeraShake Simulation of 7.7 Earthquake on the Lower San Andreas Fault

Earthquake simulations enable

- Enhanced **scientific understanding** of the physical world
- More strategic plans for bridge, building and other physical infrastructure reinforcements to **increase safety**
- Better **disaster response planning** for police, fire fighters, ER teams in high-risk areas to increase their effectiveness



Simulation courtesy of Amit Chourasia, SDSC, Table information courtesy of Southern California Earthquake Center

More Accurate Simulations Require More Infrastructure

Earthquake simulations enable

- Enhanced **scientific understanding** of the physical world
- More strategic plans for bridge, building and other physical infrastructure reinforcements to **increase safety**
- Better **disaster response planning** for police, fire fighters, ER teams in high-risk areas to increase their effectiveness

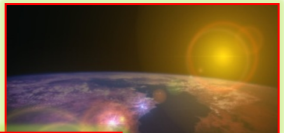
Estimated figures for simulated 240 second period, 100 hour run-time	**TERASCALE:** TeraShake domain (600x300x80 km^3)	**PETASCALE:** PetaShake domain (800x400x100 km^3)
Fault system interaction	NO	YES
Inner Scale	200m	25m
Resolution of terrain grid	1.8 billion mesh points	2.0 trillion mesh points
Magnitude of Earthquake	7.7	8.1
Time steps	20,000 (.012 sec/step)	160,000 (.0015 sec/step)
Surface data	**1.1 TB**	**1.2 PB**
Volume data	**43 TB**	**4.9 PB**

Simulation courtesy of Amit Chourasia, SDSC, Table information courtesy of Southern California Earthquake Center

Integrated Infrastructure Critical:

Application Needs Span the Spectrum

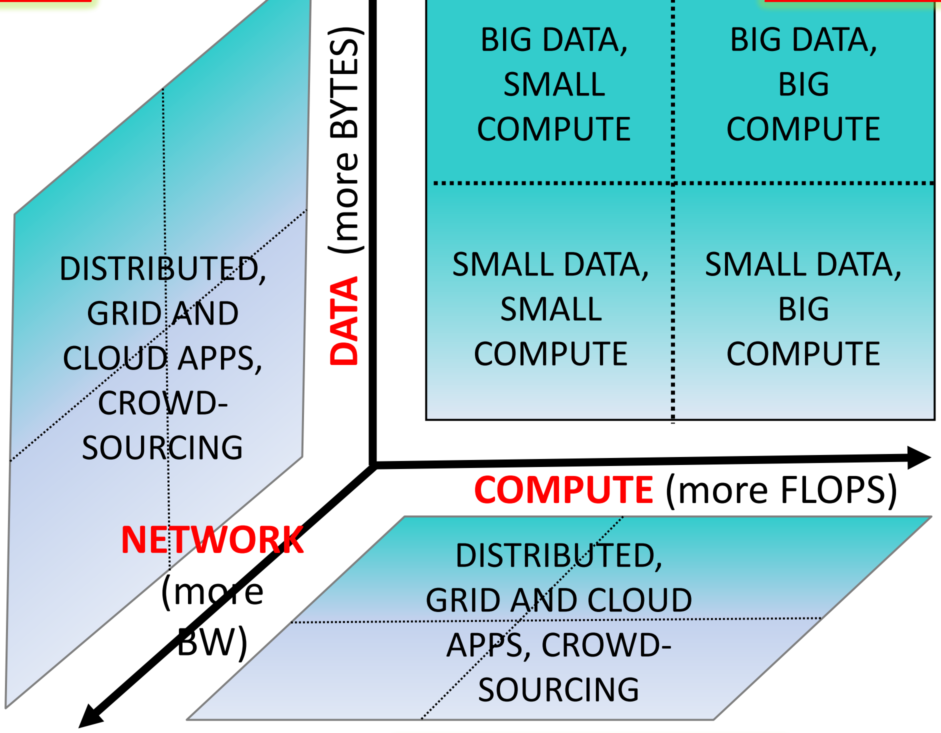
Digital history image integration



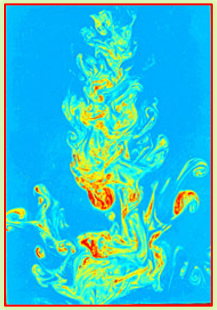
Earthquake Simulation

Cosmology

Analysis and modeling of protein function and structures

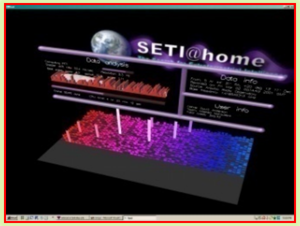


DATA (more BYTES)
BIG DATA, SMALL COMPUTE
BIG DATA, BIG COMPUTE
SMALL DATA, SMALL COMPUTE
SMALL DATA, BIG COMPUTE
COMPUTE (more FLOPS)
DISTRIBUTED, GRID AND CLOUD APPS, CROWD-SOURCING
NETWORK (more BW)
DISTRIBUTED, GRID AND CLOUD APPS, CROWD-SOURCING



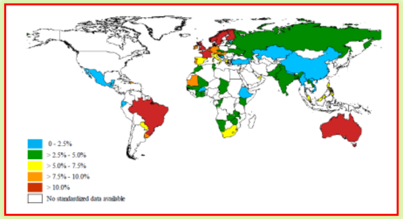
Turbulence

Analysis of Data from the CERN Large Hadron Collider

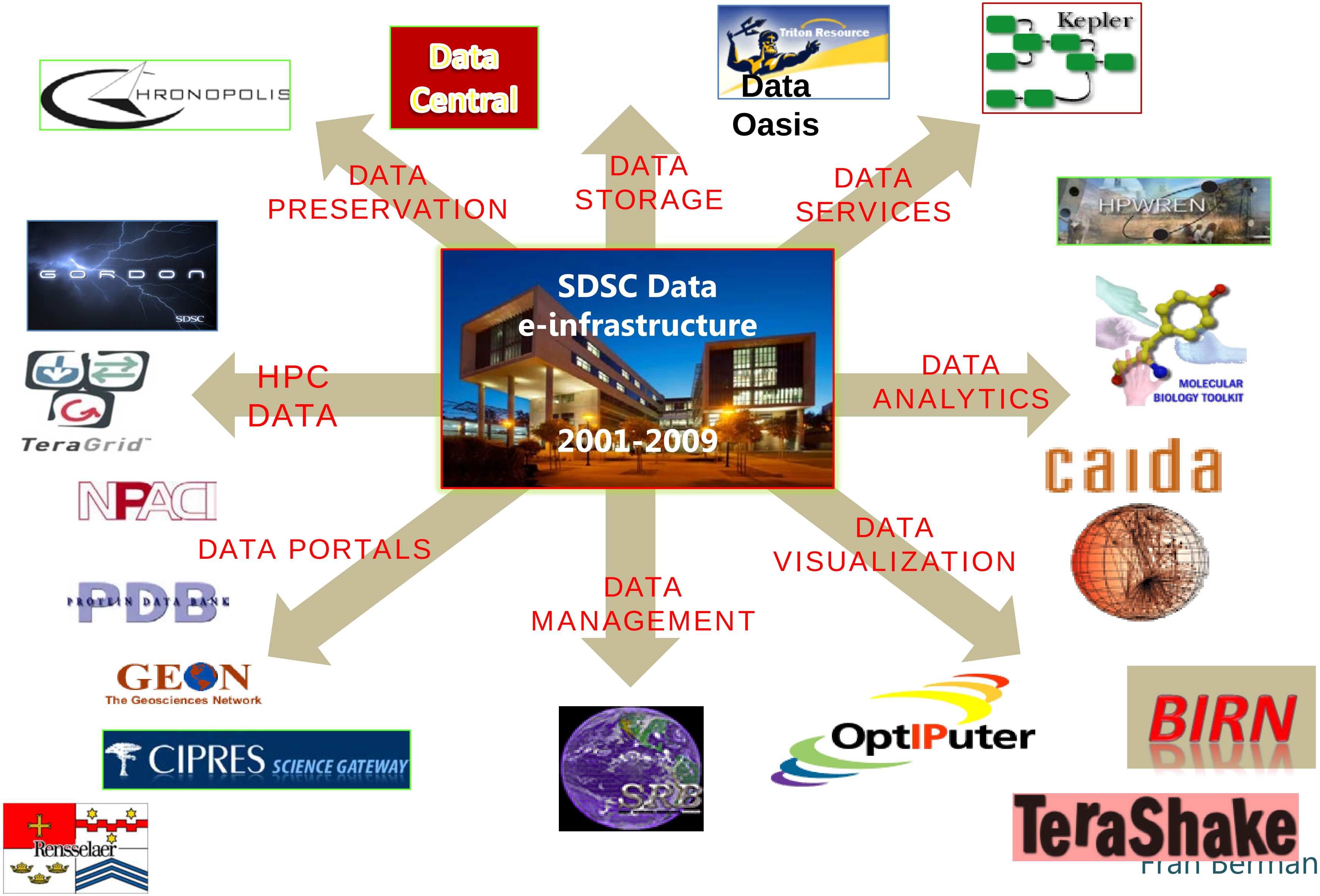


Seti@home, MilkyWay@Home, BOINC

Global Public health



Many Kinds of Technical Data Infrastructure Needed to Drive Innovation



SDSC Data e-infrastructure

2001-2009

DATA PRESERVATION

DATA STORAGE

DATA SERVICES

HPC DATA

DATA ANALYTICS

DATA PORTALS

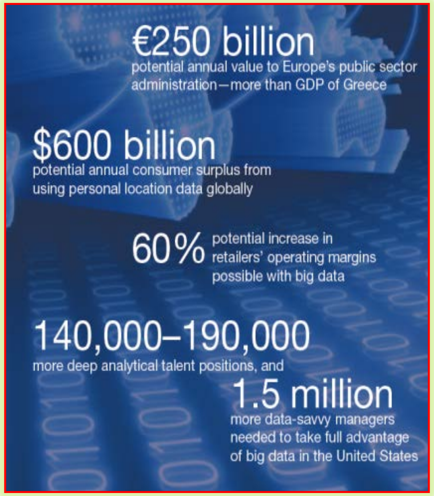
DATA MANAGEMENT

DATA VISUALIZATION

Fran Berman

Social, Organizational, and Human Infrastructure Equally Important

Human Infrastructure / Workforce



€250 billion
potential annual value to Europe's public sector administration—more than GDP of Greece

$600 billion
potential annual consumer surplus from using personal location data globally

60% potential increase in retailers' operating margins possible with big data

140,000–190,000
more deep analytical talent positions, and

1.5 million
more data-savvy managers needed to take full advantage of big data in the United States

Data Scientists

Data-focused Curriculum and Training

Social and Organizational Infrastructure

Community Practice

Sustainable Economics

RULES FOR USE OF SCHOOL GROUNDS

1. GROUNDS CLOSED FROM DUSK TO DAWN AND ALL DAY SUNDAY.
2. GROUNDS FOR USE OF ELEMENTARY AND SECONDARY SCHOOL PUPILS ONLY.
3. OTHER GROUPS BY SPECIAL BOARD PERMISSION ONLY.
4. NO LOITERING, PICNICKING OR UNLAWFUL ACTIVITY PERMITTED.
5. ALCOHOLIC BEVERAGES STRICTLY PROHIBITED.
6. TREPASSERS AND VIOLATORS SUBJECT TO ARREST AND PROSECUTION.

Policy

Common Standards

Rensselaer

McKinsey Global Institute 2011 Report,
Traffic Image: Mike Gonzalez

Today's Presentation: Two perspectives on Research Data Infrastructure

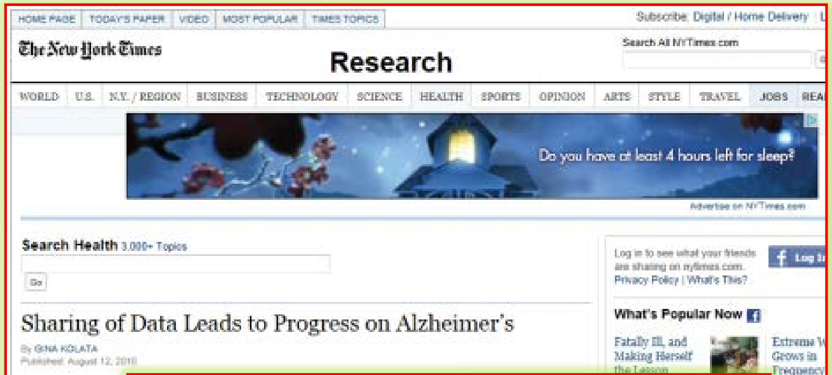
- **Opportunity: Maximizing data-driven innovation through data sharing and exchange**
  - Efforts to build a coordinated global Infrastructure to support data access, sharing and use

- **Challenge: Prioritizing the Development, Implementation and Sustainable Support of Data Infrastructure**
  - Strategies to accelerate efforts within organizations, communities and sectors

Opportunity: Maximizing Data-Driven Innovation Through Data Sharing and Exchange

Fran Berman

Data-Sharing Driving innovation Across Sectors and Communities



HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS

Subscribe: Digital / Home Delivery

The New York Times

Research

Search All NYTimes.com

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION | ARTS | STYLE | TRAVEL | JOBS | REA

Do you have at least 4 hours left for sleep?

Advertise on NYTimes.com

Search Health 3,000+ Topics

Go

Log in to see what your friends are sharing on nytimes.com. Privacy Policy | What's This?

Log In

What's Popular Now

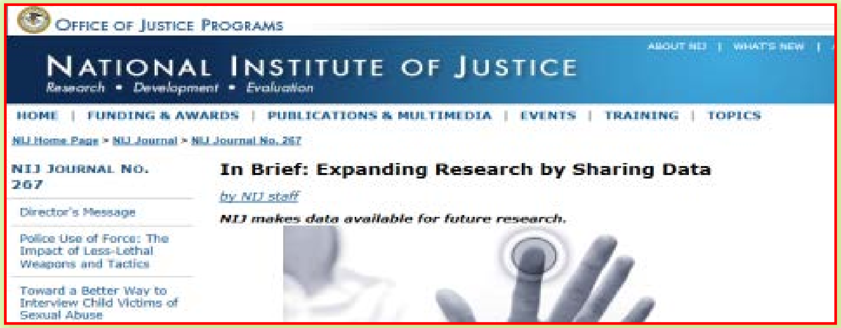
Fatally Ill, and Making Herself the Lesson

Extreme W Grows in Frequency

Sharing of Data Leads to Progress on Alzheimer's

By GINA KOLATA

Published: August 12, 2010



Office of Justice Programs

ABOUT NIJ | WHAT'S NEW |

National Institute of Justice

Research • Development • Evaluation

HOME | FUNDING & AWARDS | PUBLICATIONS & MULTIMEDIA | EVENTS | TRAINING | TOPICS

NIJ Home Page > NIJ Journal > NIJ Journal No. 267

NIJ Journal No. 267

Director's Message

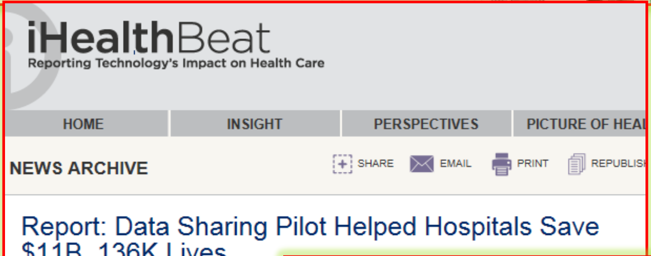
Police Use of Force: The Impact of Less-Lethal Weapons and Tactics

Toward a Better Way to Interview Child Victims of Sexual Abuse

In Brief: Expanding Research by Sharing Data

by NIJ staff

NIJ makes data available for future research.



iHealthBeat

Reporting Technology's Impact on Health Care

HOME | INSIGHT | PERSPECTIVES | PICTURE OF HEAL

NEWS ARCHIVE

SHARE | EMAIL | PRINT | REPUBLISH

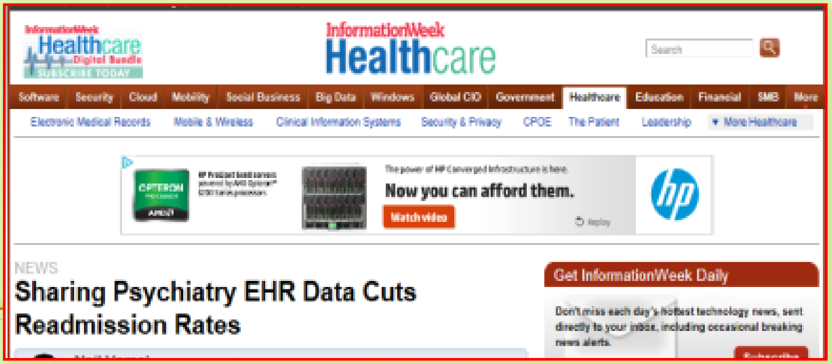
Report: Data Sharing Pilot Helped Hospitals Save $11B, 136K Lives

Tuesday, February 18, 2014

TOPIC ALERT:

- Big Data

A quality im hospitals sa five-and-a-h QUEST col (Thomas, C



InformationWeek Healthcare Digital Bundle SUBSCRIBE TODAY

InformationWeek Healthcare

Search

Software | Security | Cloud | Mobility | Social Business | Big Data | Windows | Global CIO | Government | Healthcare | Education | Financial | SMB | More

Electronic Medical Records | Mobile & Wireless | Clinical Information Systems | Security & Privacy | CPOE | The Patient | Leadership | More Healthcare

The power of HP Converged Infrastructure is here.

Now you can afford them.

Watch video

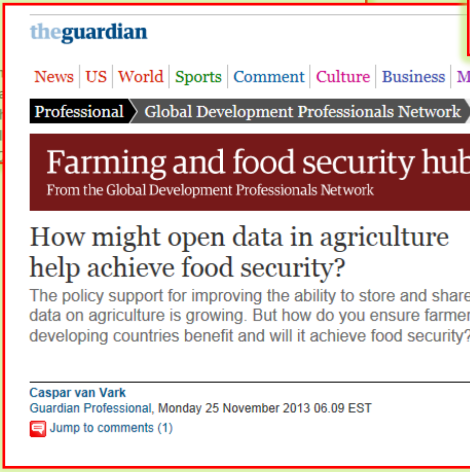
Replay

NEWS

Sharing Psychiatry EHR Data Cuts Readmission Rates

Get InformationWeek Daily

Don't miss each day's hottest technology news, sent directly to your inbox, including occasional breaking news alerts.



theguardian

News | US | World | Sports | Comment | Culture | Business | Money

Professional | Global Development Professionals Network | Sign

Farming and food security hub

From the Global Development Professionals Network

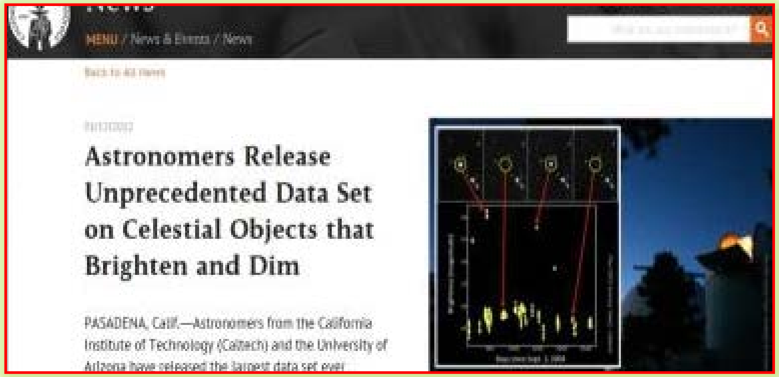
How might open data in agriculture help achieve food security?

The policy support for improving the ability to store and share data on agriculture is growing. But how do you ensure farmers in developing countries benefit and will it achieve food security?

Caspar van Vark

Guardian Professional, Monday 25 November 2013 06.09 EST

Jump to comments (1)



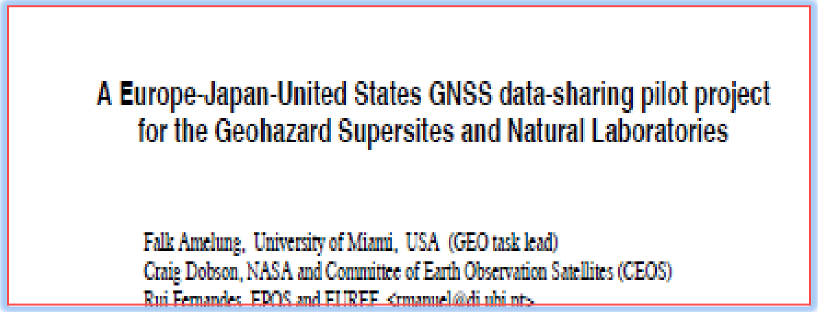
MENU / News & Events / News

Back to All News

Astronomers Release Unprecedented Data Set on Celestial Objects that Brighten and Dim

PASADENA, Calif.—Astronomers from the California Institute of Technology (Caltech) and the University of Arizona have released the largest data set ever

World-wide Efforts Focusing on Infrastructure to Support Research Data Sharing, Access, Use



A Europe-Japan-United States GNSS data-sharing pilot project for the Geohazard Supersites and Natural Laboratories

Falk Amelung, University of Miami, USA (GEO task lead)
Craig Dobson, NASA and Committee of Earth Observation Satellites (CEOS)
Rui Fernandes, EPOS and EUREF <rmanuel@di.ubi.pt>

SCIENCE BUSINESS

Bringing tog
Research | I

Policy Analysis: Investment, Policy

Horizon 2020: Special Coverage

SPECIAL COVERAGE HORIZON 2020

PUBLISHED: 18 SEPTEMBER 2013

Promote data sharing to advance global research say policy leaders

Richard L. Hudson, Science|Business

EU and US experts see big benefits from scientists sharing more data - but say global agreement on privacy, literacy and other issues is needed

Related News

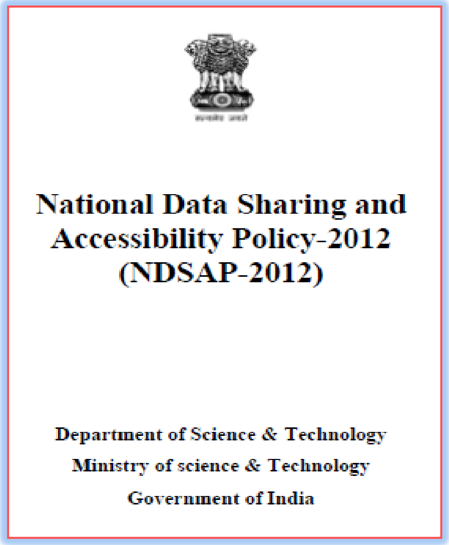
'This is not the time for change or delay'
25 September 2013

Science, Humanities, Arts Communities

DATA.GOV
EMPOWERING PEOPLE

Libraries, Archives, Repositories, Museums

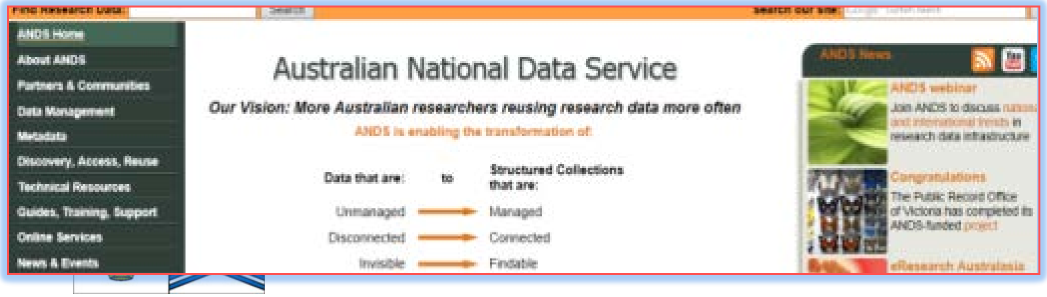
E-Infrastructure professionals, data analysts, data center staff, …



National Data Sharing and Accessibility Policy-2012 (NDSAP-2012)

Department of Science & Technology
Ministry of science & Technology
Government of India

Data Scientists



Australian National Data Service

Our Vision: More Australian researchers reusing research data more often

ANDS is enabling the transformation of:

Data that are:	to	Structured Collections that are:
Unmanaged	→	Managed
Disconnected	→	Connected
Invisible	→	Findable

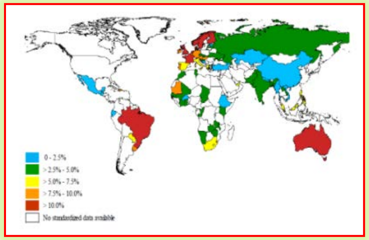
ANDS webinar
Join ANDS to discuss national and international trends in research data infrastructure

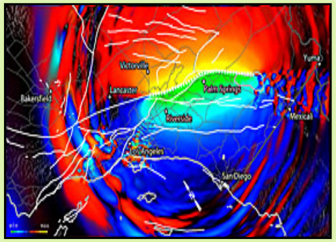
Congratulations
The Public Record Office of Victoria has completed its ANDS-funded project

eResearch Australasia

Fran Berman

Many Infrastructure Building Blocks Needed to Accelerate Progress





Research Dissemination and Reproducibility

Data Use and Re-use

Data Access (now) and Preservation (later)

Data Discovery and Data Sharing

Data Access and Distribution Policy

Institutional Data Sharing Practice

Data Discovery Tools

Digital Object Identifiers

Common Metadata Standards

Data Citation Standards

Data Preservation Practice

Data Analytics Algorithms

Data Scientists and Expert Support

Curation Practice and Policy

Sustainable Economic Models

Auditing, Certification and Reporting Practice

Fran Berman

Fran Berman

Research Data Alliance Created to Accelerate Development of Research Data Sharing Infrastructure Worldwide

- RDA is an emerging, global community-driven organization created to **accelerate the development of research data-sharing infrastructure** world-wide.
- RDA community efforts focus on building **social, organizational and technical infrastructure** to
  - *reduce barriers to data sharing and exchange*
  - *accelerate the development of coordinated global data infrastructure*

RDA Approach:
CREATE → ADOPT → USE

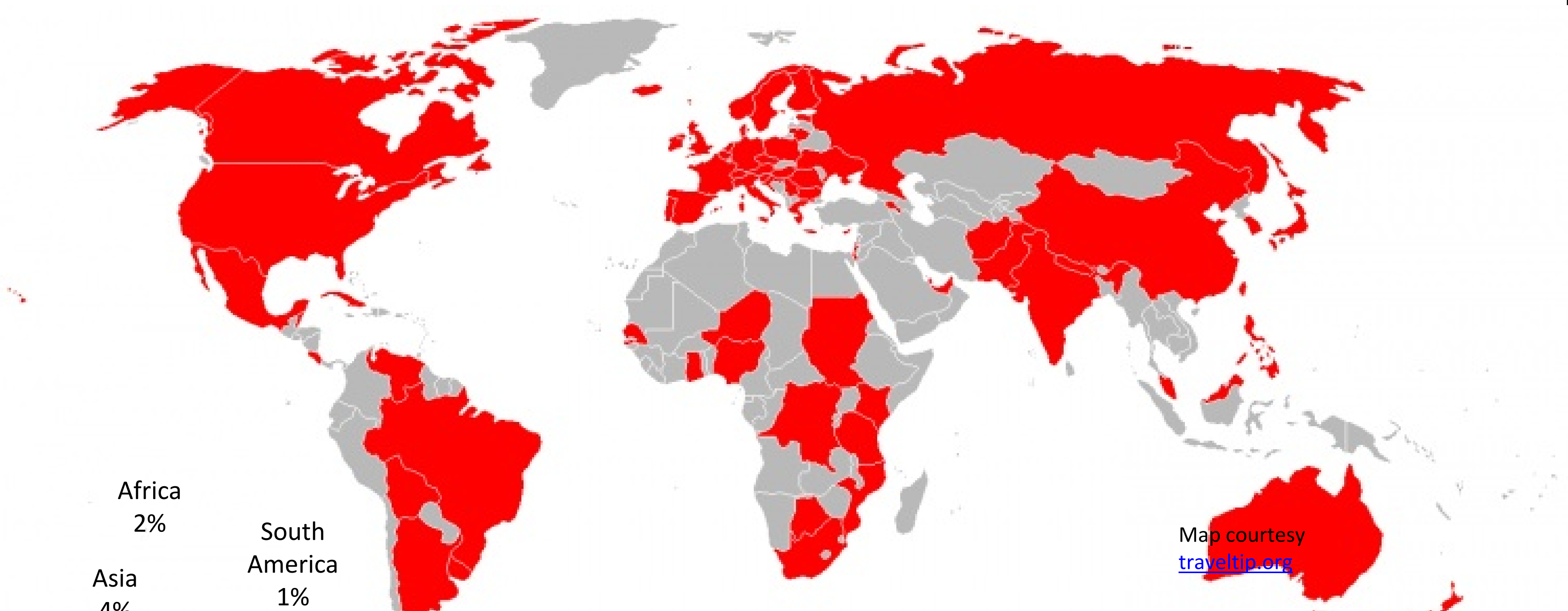
RDA Members come together as

- **Working Groups** – 12-18 month efforts to build, adopt, and use specific pieces of infrastructure
- **Interest Groups** – longer-lived discussion forums that spawn Working Groups as specific pieces of needed infrastructure are identified.

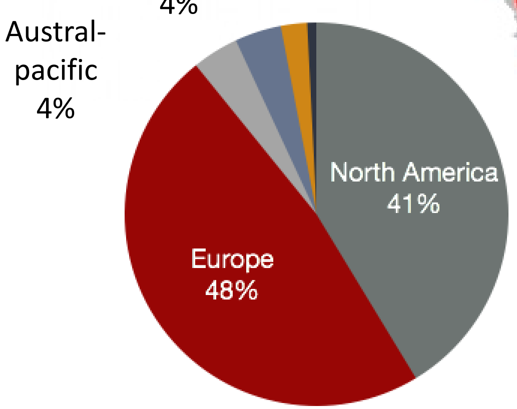
Working Group efforts focus on the development and use of data sharing infrastructure

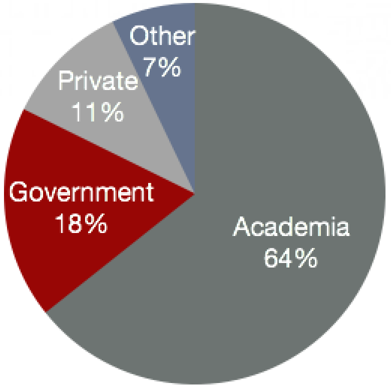
- **Code, policy, infrastructure, standards, or best practices that are adopted and used** by communities to enable data sharing
- **"Harvestable" efforts** for which 12-18 months of work can eliminate a roadblock
- **Efforts that have substantive applicability** to groups within the data community, but may not apply to everyone
- **Efforts for which working scientists and researchers can start today**

The RDA Community Today: Over 1600 members from 70+ countries (as of 15/3/14)



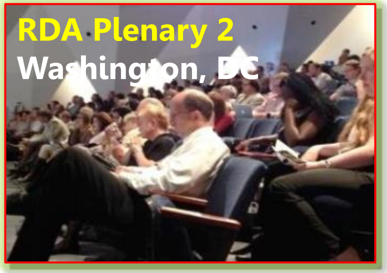
Map courtesy traveltip.org





Precipitous Growth

RDA Plenary 1 / Launch
Gothenburg, Sweden


RDA Plenary 2
Washington, DC

RDA Plenary 3
Dublin, Ireland

Global Data Planning Meeting: October 2012

First RDA organizational telecon: August 2012

First Working Groups and Interest Groups

240 participants

RDA Launch / First Plenary

March 2013

First "neutral space" community meeting (Data Citation Summit)

First Org. Partner Meet-up

First BOFs

380 participants from 22 countries

RDA Second Plenary

September 2013

First Organizational Assembly

6 co-located events

14 BOF,
12 Working Groups, 22 Interest Groups

497 participants

RDA Third Plenary

March 2014

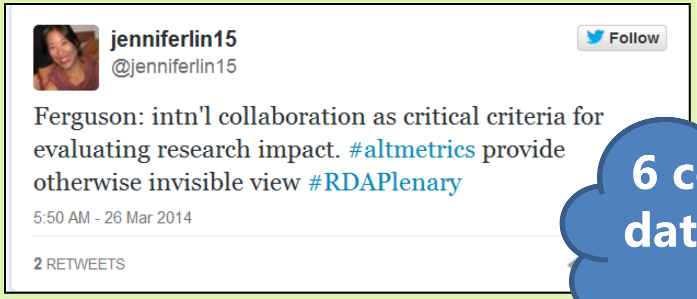
First Working Group exchange meeting

Amsterdam

RDA Fourth Plenary

September 2014

RDA Plenary 3
March 26-28, 2014
Dublin, Ireland



jenniferlin15
@jenniferlin15

Ferguson: intn'l collaboration as critical criteria for evaluating research impact. #altmetrics provide otherwise invisible view #RDAPlenary

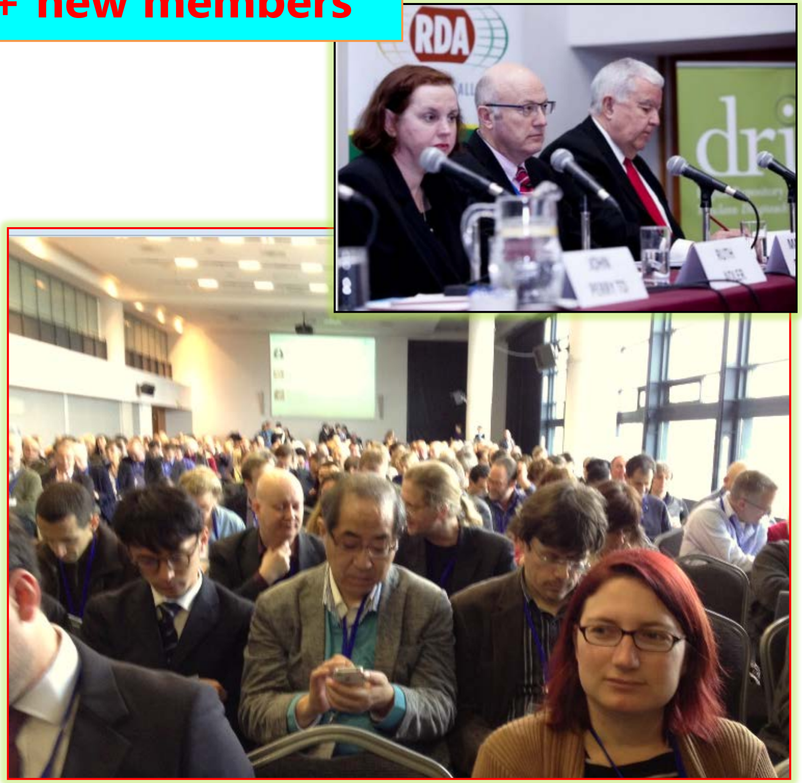
5:50 AM - 26 Mar 2014

2 RETWEETS

6 co-located data-focused events

Organizational Assembly Meeting: 20+ new members

Professional Title	Total	%
Advisor/Consultant	22	4%
CEO / Managing Director / Chief Executive	35	7%
CTO / IT Director	20	4%
IT Specialist / IT Architect	53	11%
Journalist / Editor / Copywriter	6	1%
Librarian	27	5%
Other	93	19%
Student	38	8%
Policy Development Manager / Policy Consultant	12	2%
Professor	42	8%
Programme Manager / Project Manager	62	12%
Researcher	87	18%
Total	497	100%



Fran Berman

RDA Interest (IG) and Working Groups (WG) by Focus

(as of 15/3/14)

Domain Science - focused

- Toxicogenomics Interoperability IG
- Structural Biology IG
- Biodiversity Data Integration IG
- Agricultural Data Interoperability IG
- Digital History and Ethnography IG
- Defining Urban Data Exchange for Science IG
- Marine Data Harmonization IG
- Materials Data Management IG

Community Needs - focused

- Community Capability Model IG
- Engagement IG
- Clouds in Developing Countries IG

Reference and Sharing - focused

- Data Citation IG
- Data Categories and Codes WG
- Legal Interoperability IG

Data Stewardship - focused

- Research Data Provenance IG
- Certification of Digital Repositories IG
- Preservation e-infrastructure
- Long-tail of Research Data IG
- Publishing Data IG
- Domain Repositories IG
- Global Registry of Trusted Data Repositories and Services IG

Base Infrastructure - focused

- Data Foundations and Terminology WG
- Metadata Standards WG
- Practical Policy WG
- PID Information Types WG
- Data Type Registries WG
- Metadata IG
- Big Data Analytics IG
- Data Brokering IG

HPC Members welcome!

First RDA Infrastructure Deliverables coming this Fall

Data Type Registries WG

- **Deliverables:** System of data type registries, formal model for describing types, working model of a registry.
- **Initial Adopters and Users:** CNRI, International DOI Foundation, Deep Carbon Observatory

Practical Code Policies

- **Deliverables:** Survey of policies in production use, testbed of machine actionable policies, deployment of 5 policy sets, policy starter kits
- **Initial Adopters and Users:** RENCI, DataNet Federation Consortium, CESNET, Odum Institute, EUDAT

Persistent Identifier Information Types

- **Deliverables:** Minimal set of PID types, API
- **Initial Adopters and Users:** Data Conservancy, DKRZ

Language Codes

- **Deliverables:** Operationalization of ISO language categories for repositories.
- **Initial Adopters and Users:** Language Archive, Paradisec

Data Foundations and Terminology

- **Deliverables:** Common vocabulary for data terms, formal definitions and open registry for data terms
- **Initial Adopters and Users:** EUDAT, DKRZ, Deep Carbon Observatory, CLARIN, EPOS

Metadata Standards

- **Deliverables:** Use cases and prototype directory of current metadata standards starting from DCC directory
- **Initial Adopters and Users:** JISC, DataOne

Next Steps for the RDA

More Infrastructure

Continuing pipeline of infrastructure deliverables adopted and used to accelerate data sharing

Increasing coordination of infrastructure

Effective Community

Increasing cross-boundary collaborations between domains, sectors, organizations

Synergistic Programs

International and regional programs focusing on workforce, outreach, expansion of infrastructure impact

Focus on Industry

New partners in the Organizational Assembly

Focused strategy to support development of industry infrastructure for data sharing

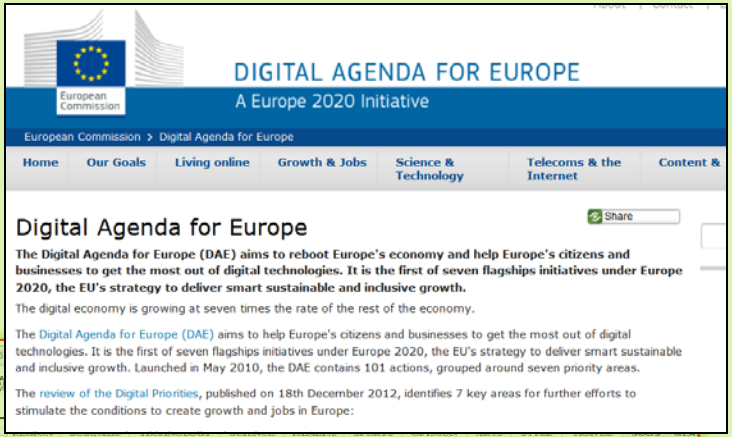
Challenge: Supporting and Sustaining Research Data Infrastructure

Fran Berman

Research Data and Data Sharing Key to Innovation.
Research Data an Accelerator for All Sectors.

- National governments increasingly calling for public access to research data.
  - *Valuable to all sectors as a driver of current and future innovation*

- **Research data infrastructure is necessary** to support
  - *Use and re-use*
  - *Research reproducibility*
  - *Federally mandated data management plans*
  - *Public access to research data*


DIGITAL AGENDA FOR EUROPE

A Europe 2020 Initiative

European Commission > Digital Agenda for Europe

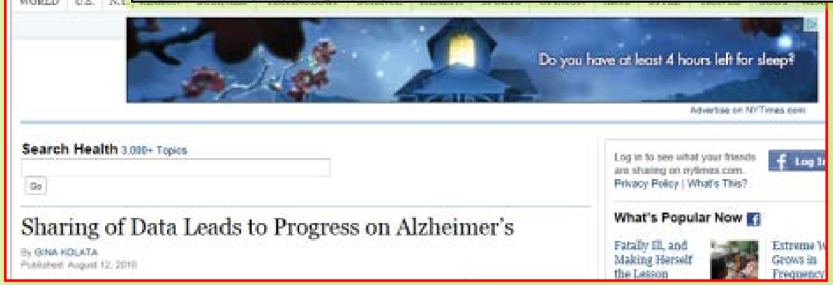
Digital Agenda for Europe

The Digital Agenda for Europe (DAE) aims to reboot Europe's economy and help Europe's citizens and businesses to get the most out of digital technologies. It is the first of seven flagships initiatives under Europe 2020, the EU's strategy to deliver smart sustainable and inclusive growth.

The digital economy is growing at seven times the rate of the rest of the economy.

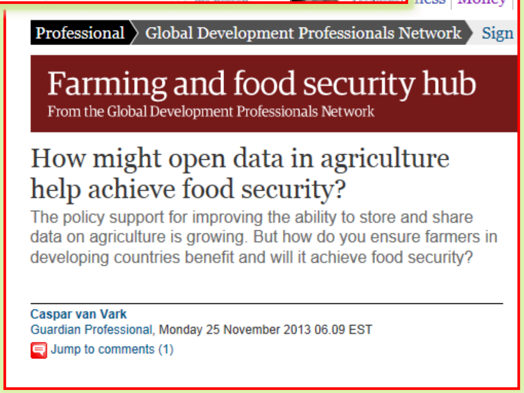
The Digital Agenda for Europe (DAE) aims to help Europe's citizens and businesses to get the most out of digital technologies. It is the first of seven flagships initiatives under Europe 2020, the EU's strategy to deliver smart sustainable and inclusive growth. Launched in May 2010, the DAE contains 101 actions, grouped around seven priority areas.

The review of the Digital Priorities, published on 18th December 2012, identifies 7 key areas for further efforts to stimulate the conditions to create growth and jobs in Europe:


HOME PAGE
TODAY'S
The New York T
WORLD
U.S.
N.Y.
Do you have at least 4 hours left for sleep?
Advertise on NYTimes.com
Search Health 3,000+ Topics
Go
Log in to see what your friends are sharing on nytimes.com.
Privacy Policy | What's This?
Log In
What's Popular Now
Fatally Ill, and Making Herself the Lesson
Extreme W Grows in Frequency

Sharing of Data Leads to Progress on Alzheimer's

By GINA KOLATA
Published: August 12, 2010



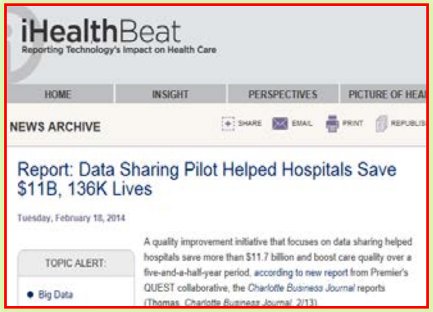
Farming and food security hub

From the Global Development Professionals Network

How might open data in agriculture help achieve food security?

The policy support for improving the ability to store and share data on agriculture is growing. But how do you ensure farmers in developing countries benefit and will it achieve food security?

Caspar van Vark
Guardian Professional, Monday 25 November 2013 06.09 EST
Jump to comments (1)



NEWS ARCHIVE

Report: Data Sharing Pilot Helped Hospitals Save $11B, 136K Lives

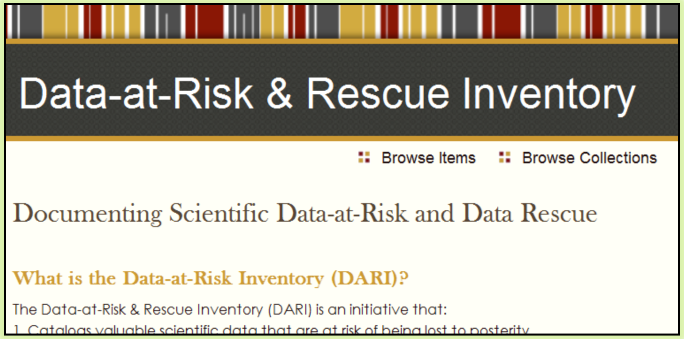
Tuesday, February 18, 2014

A quality improvement initiative that focuses on data sharing helped hospitals save more than $11.7 billion and boost care quality over a five-and-a-half-year period, according to new report from Premier's QUEST collaborative, the *Charlotte Business Journal* reports (Thomas, *Charlotte Business Journal*, 2/13)

Rensselaer

Fran Berman

Yet Research Data Infrastructure is Difficult to Sustain. Why?



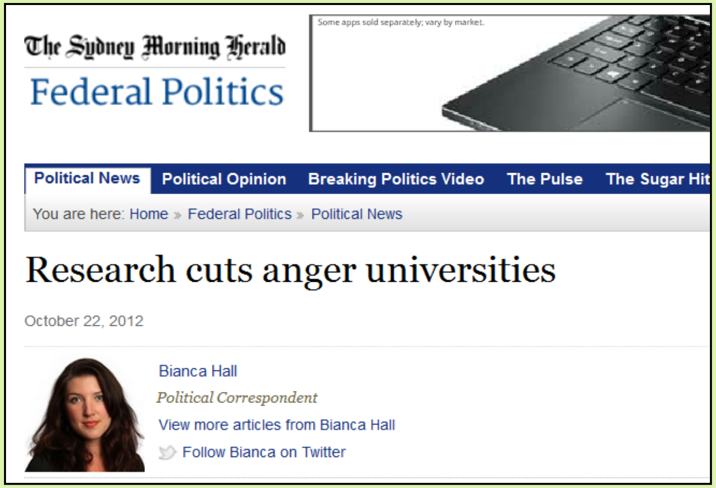
Data-at-Risk & Rescue Inventory

Documenting Scientific Data-at-Risk and Data Rescue

What is the Data-at-Risk Inventory (DARI)?

The Data-at-Risk & Rescue Inventory (DARI) is an initiative that:

1. Catalogs valuable scientific data that are at risk of being lost to posterity



The Sydney Morning Herald

Federal Politics

You are here: Home » Federal Politics » Political News

Research cuts anger universities

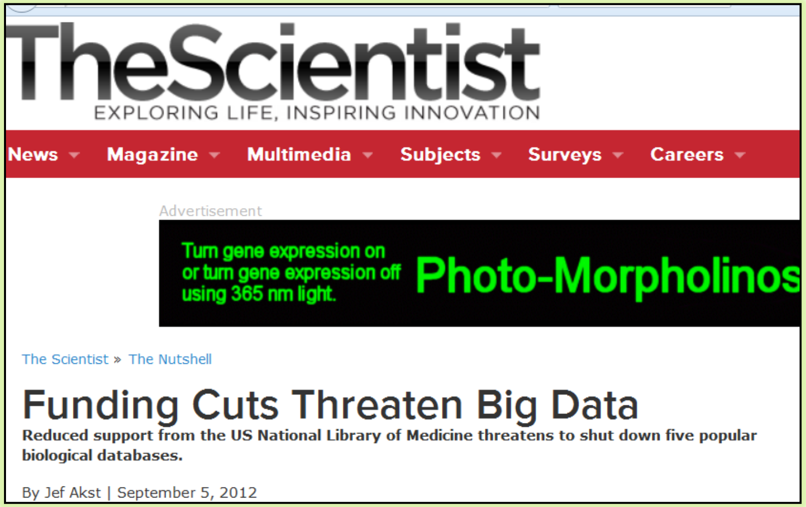
October 22, 2012

Bianca Hall

Political Correspondent

View more articles from Bianca Hall

Follow Bianca on Twitter



TheScientist

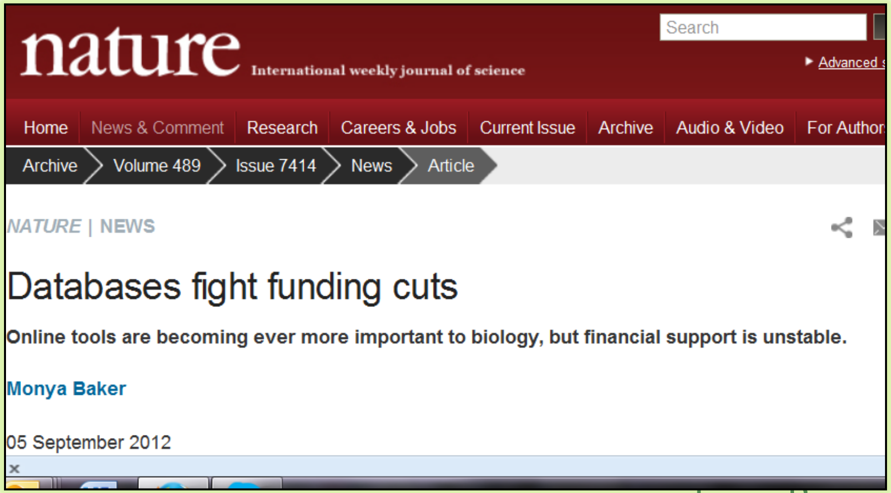
EXPLORING LIFE, INSPIRING INNOVATION

The Scientist » The Nutshell

Funding Cuts Threaten Big Data

Reduced support from the US National Library of Medicine threatens to shut down five popular biological databases.

By Jef Akst | September 5, 2012



nature International weekly journal of science

NATURE | NEWS

Databases fight funding cuts

Online tools are becoming ever more important to biology, but financial support is unstable.

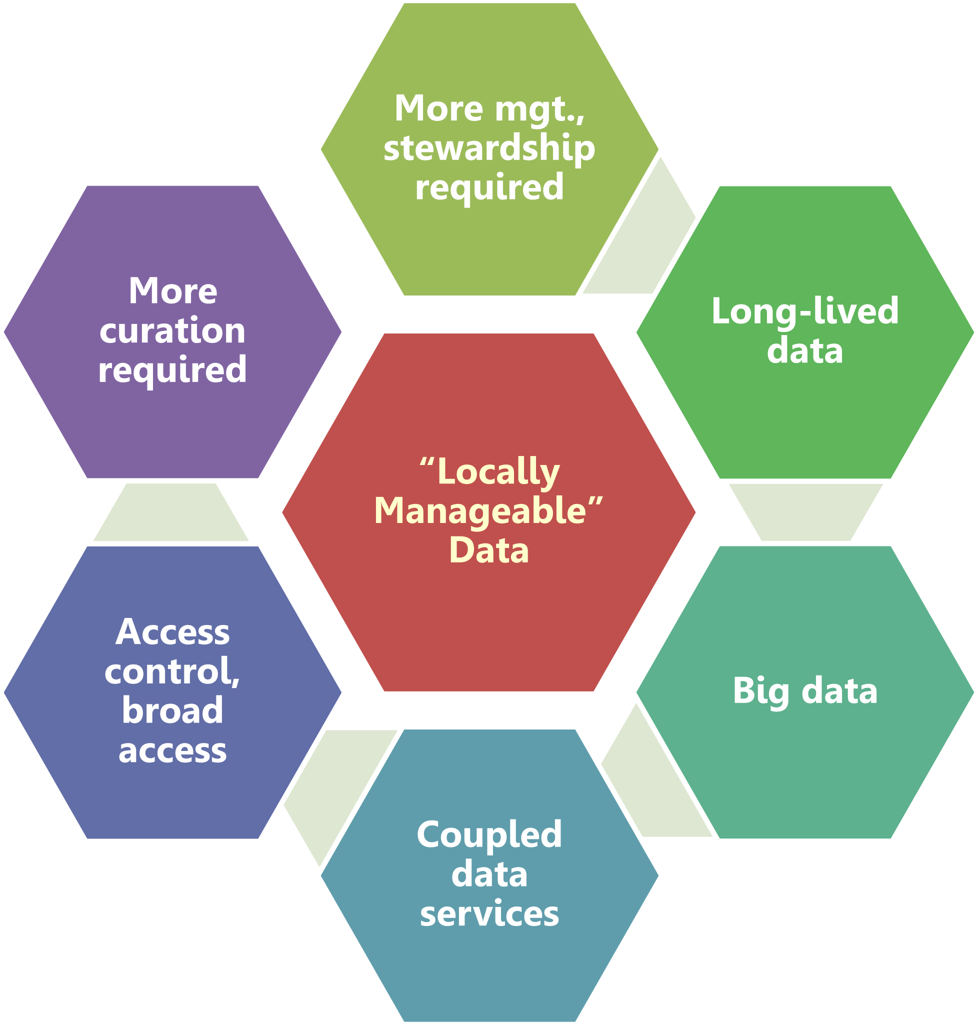
Monya Baker

05 September 2012

Sustainable Data Infrastructure Starts with a Sustainable Economic Model

It's not just about the cost of storage. Research data infrastructure costs increase with usage, stewardship and access requirements, perceived value

Greater costs beyond the center (including "big" data)


More mgt., stewardship required
More curation required
Long-lived data
"Locally Manageable" Data
Access control, broad access
Big data
Coupled data services

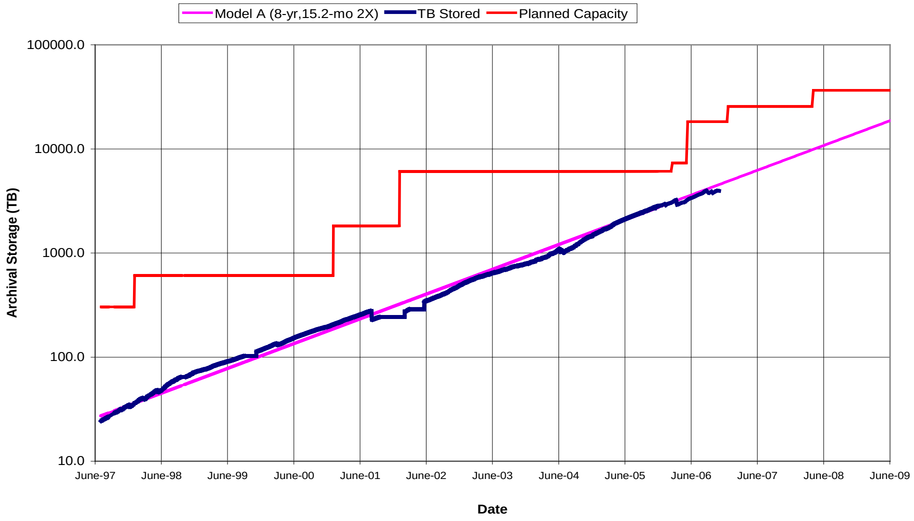
Fran Berman

Data Economics: Data Management, Stewardship, and Use Incur Continuing Infrastructure Costs

Costs include

- Maintenance and upkeep
- Software tools and packages
- Utilities (power, cooling)
- Space
- Networking
- Security and failover systems
- People (expertise, help, infrastructure management, development)
- Training, documentation
- Monitoring, auditing
- Reporting costs
- Costs of compliance with regulation, etc.

Resources and Resource Refresh



SDSC Data Storage Growth '97-'09

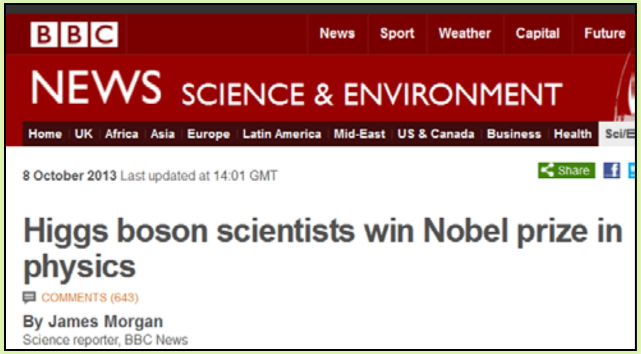
- *Most valuable data replicated*
- *As research collections increase, storage capacity must stay ahead of demand*

Information courtesy of Richard Moore, SDSC

In the Public Sector, Research and Infrastructure often compete for limited funding.

Infrastructure Investment a hard sell …

	Research	Infrastructure
What is Newsworthy?	New discoveries and breakthroughs	Failure of systems
What is the value proposition?	Domain and national leadership and competitiveness	Enabler of innovation
Funding Model	Fixed-term funding	Continuous long-term support

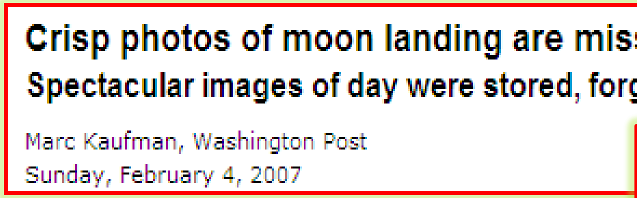


8 October 2013 Last updated at 14:01 GMT

Higgs boson scientists win Nobel prize in physics

COMMENTS (643)

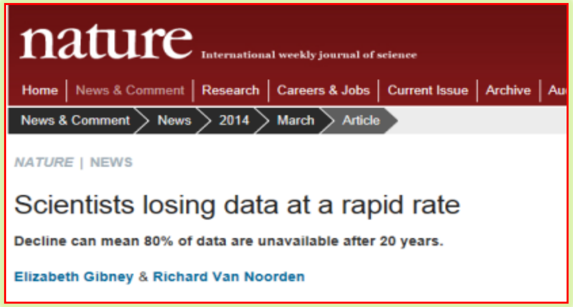
By James Morgan
Science reporter, BBC News



Crisp photos of moon landing are missing

Spectacular images of day were stored, forgotten -- and lost

Marc Kaufman, Washington Post
Sunday, February 4, 2007



NATURE | NEWS

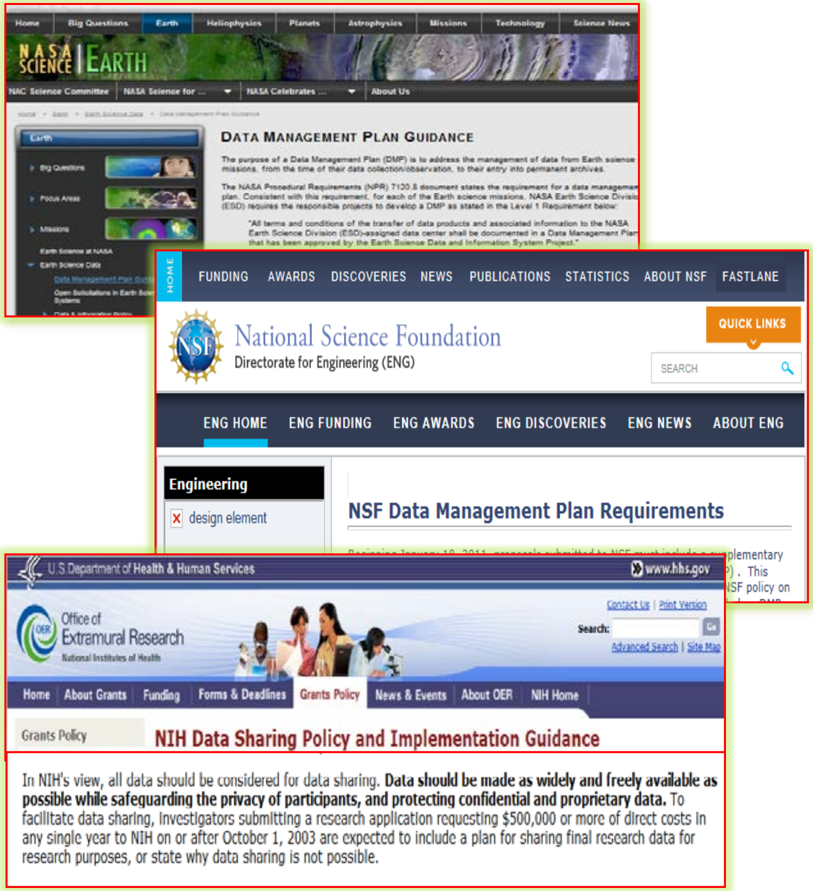
Scientists losing data at a rapid rate

Decline can mean 80% of data are unavailable after 20 years.

Elizabeth Gibney & Richard Van Noorden

Fran Berman

Data Infrastructure particularly important in light of increasing National R&D Agency Requirements for Data Access and Management



Data Management Plan Guidance

The purpose of a Data Management Plan (DMP) is to address the management of data from Earth science missions, from the time of their data collection/observation, to their entry into permanent archives.

The NASA Procedural Requirements (NPR) 7120.8 document states the requirement for a data managemen plan. Consistent with this requirement, for each of the Earth science missions, NASA Earth Science Divisio (ESD) requires the responsible projects to develop a DMP as stated in the Level 1 Requirement below:

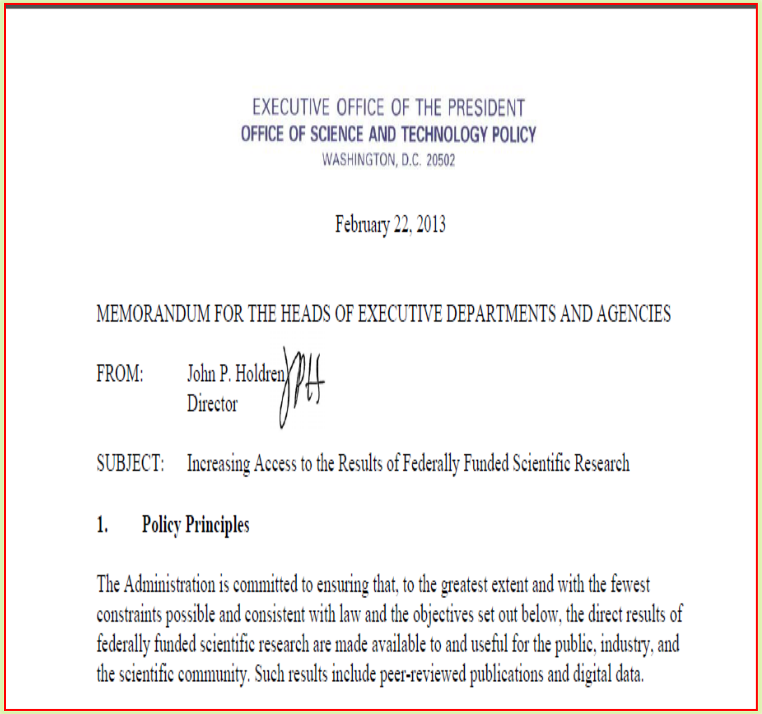
> "All terms and conditions of the transfer of data products and associated information to the NASA Earth Science Division (ESD)-assigned data center shall be documented in a Data Management Plan that has been approved by the Earth Science Data and Information System Project."

NSF Data Management Plan Requirements

Beginning January 18, 2011, proposals submitted to NSF must include a supplementary
P) . This
NSF policy on

NIH Data Sharing Policy and Implementation Guidance

In NIH's view, all data should be considered for data sharing. **Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data.** To facilitate data sharing, investigators submitting a research application requesting $500,000 or more of direct costs in any single year to NIH on or after October 1, 2003 are expected to include a plan for sharing final research data for research purposes, or state why data sharing is not possible.



EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

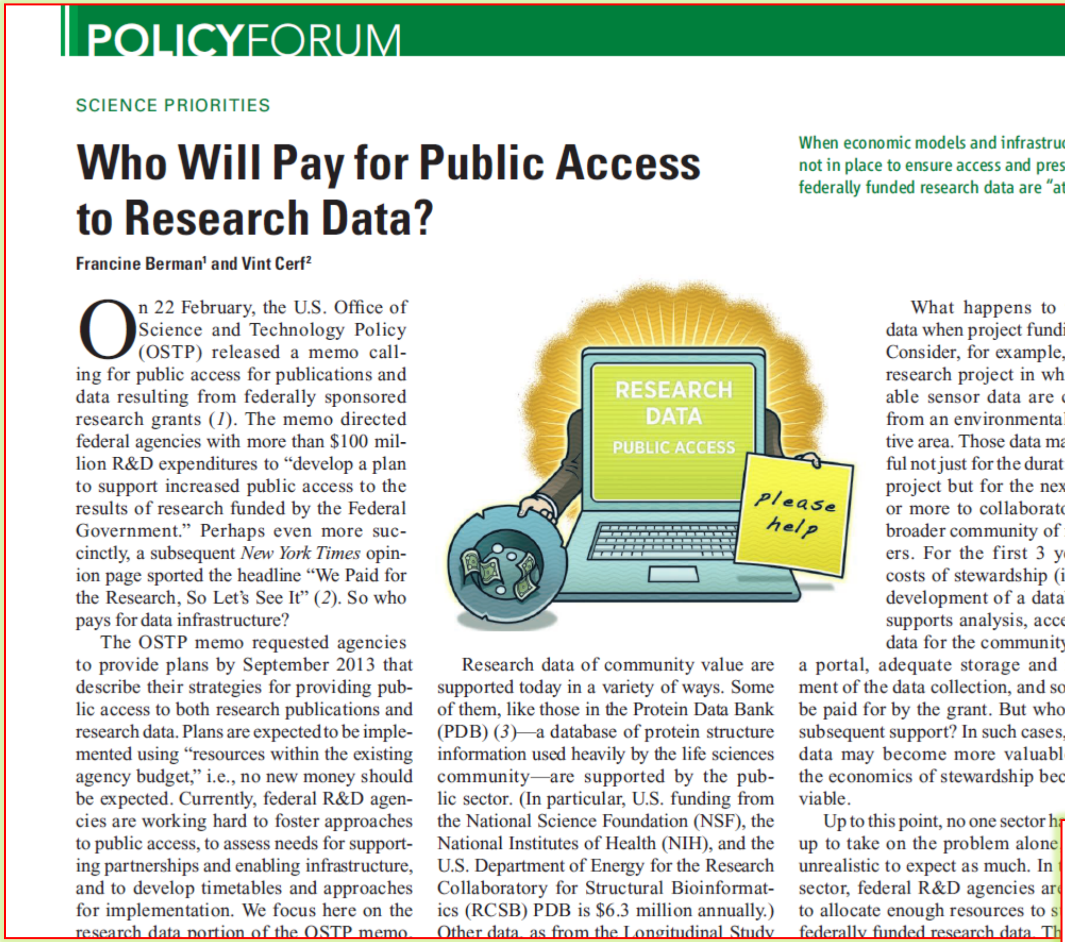
FROM: John P. Holdren JPH
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

1. **Policy Principles**

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

Economics of Public Access:
Who Pays the Data Bill?



POLICYFORUM

SCIENCE PRIORITIES

Who Will Pay for Public Access to Research Data?

When economic models and infrastructure are not in place to ensure access and preservation, federally funded research data are "at risk."

Francine Berman[1] and Vint Cerf[2]

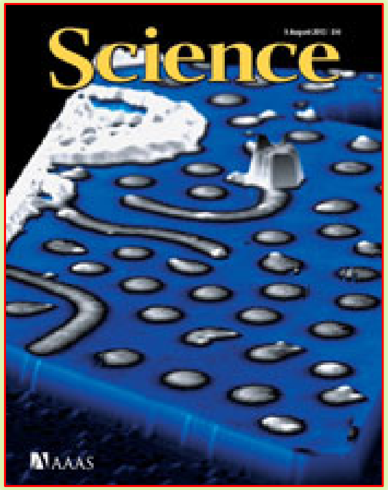
On 22 February, the U.S. Office of Science and Technology Policy (OSTP) released a memo calling for public access for publications and data resulting from federally sponsored research grants (*1*). The memo directed federal agencies with more than $100 million R&D expenditures to "develop a plan to support increased public access to the results of research funded by the Federal Government." Perhaps even more succinctly, a subsequent *New York Times* opinion page sported the headline "We Paid for the Research, So Let's See It" (*2*). So who pays for data infrastructure?

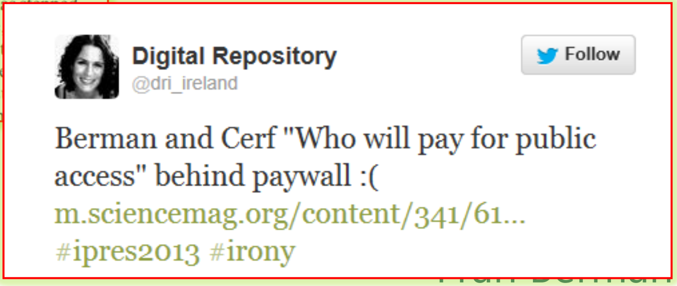
The OSTP memo requested agencies to provide plans by September 2013 that describe their strategies for providing public access to both research publications and research data. Plans are expected to be implemented using "resources within the existing agency budget," i.e., no new money should be expected. Currently, federal R&D agencies are working hard to foster approaches to public access, to assess needs for supporting partnerships and enabling infrastructure, and to develop timetables and approaches for implementation. We focus here on the research data portion of the OSTP memo.

Research data of community value are supported today in a variety of ways. Some of them, like those in the Protein Data Bank (PDB) (*3*)—a database of protein structure information used heavily by the life sciences community—are supported by the public sector. (In particular, U.S. funding from the National Science Foundation (NSF), the National Institutes of Health (NIH), and the U.S. Department of Energy for the Research Collaboratory for Structural Bioinformatics (RCSB) PDB is $6.3 million annually.) Other data, as from the Longitudinal Study

What happens to valuable data when project funding ends? Consider, for example, a 3-year research project in which valuable sensor data are collected from an environmentally sensitive area. Those data may be useful not just for the duration of the project but for the next decade or more to collaborators and a broader community of researchers. For the first 3 years, the costs of stewardship (including development of a database that supports analysis, access to the data for the community through a portal, adequate storage and management of the data collection, and so on) may be paid for by the grant. But who pays for subsequent support? In such cases, research data may become more valuable just as the economics of stewardship become less viable.

Up to this point, no one sector h up to take on the problem alone unrealistic to expect as much. In t sector, federal R&D agencies ar to allocate enough resources to s federally funded research data. Th



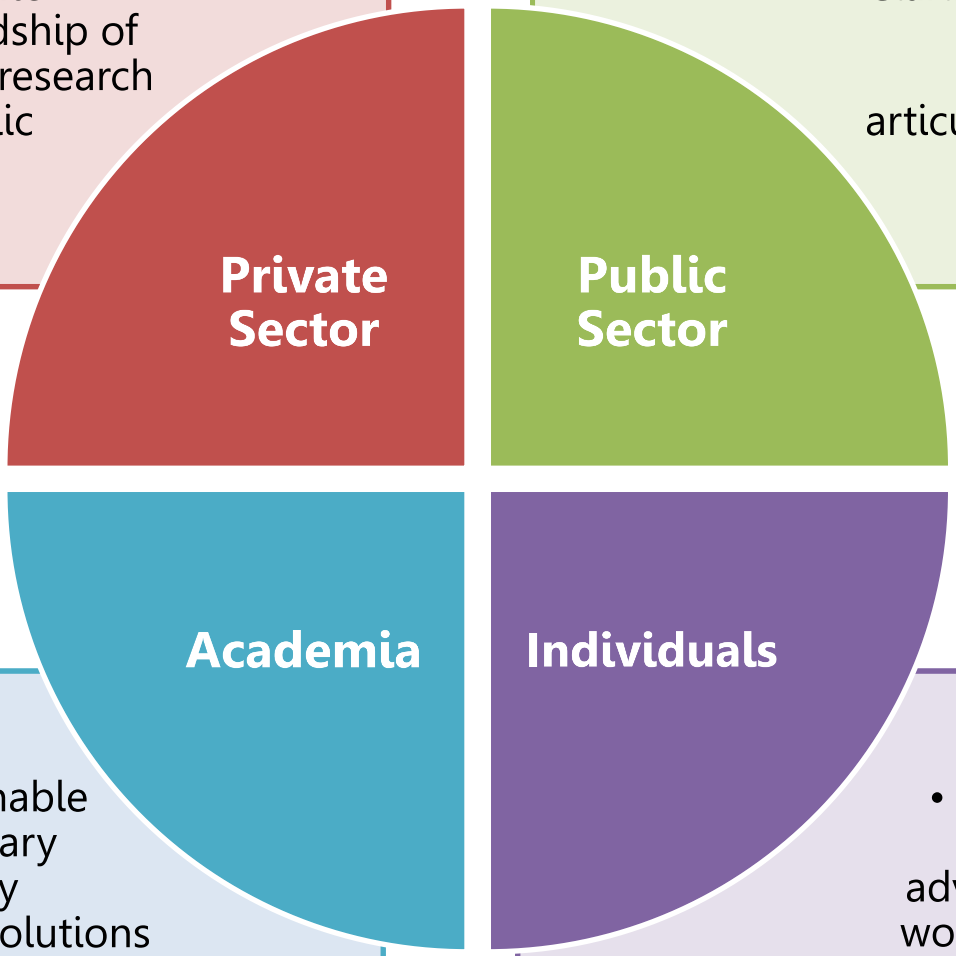


Digital Repository
@dri_ireland

Berman and Cerf "Who will pay for public access" behind paywall :(
m.sciencemag.org/content/341/61...
#ipres2013 #irony

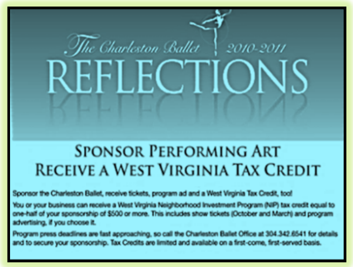
Article: Science Magazine, August 9, 2013. Free public access link at http:/www.cs.rpi.edu/~bermaf/

Op-Ed Recommendations: Partner Across Sectors to Distribute the Preservation and Stewardship Responsibilities



- Facilitate private sector stewardship of public access research data as a public good

- Clarify public sector stewardship commitments: articulate what data will / won't be supported



The Charleston Ballet 2010-2011

REFLECTIONS

SPONSOR PERFORMING ART
RECEIVE A WEST VIRGINIA TAX CREDIT

Sponsor the Charleston Ballet, receive tickets, program ad and a West Virginia Tax Credit, too!

You or your business can receive a West Virginia Neighborhood Investment Program (NIP) tax credit equal to one-half of your sponsorship of $500 or more. This includes show tickets (October and March) and program advertising, if you choose it.

Program press deadlines are fast approaching, so call the Charleston Ballet Office at 304.342.6541 for details and to secure your sponsorship. Tax Credits are limited and available on a first-come, first-served basis.

- Create sustainable university library and repository stewardship solutions

- Evolve research culture to take advantage of what works in the private sector

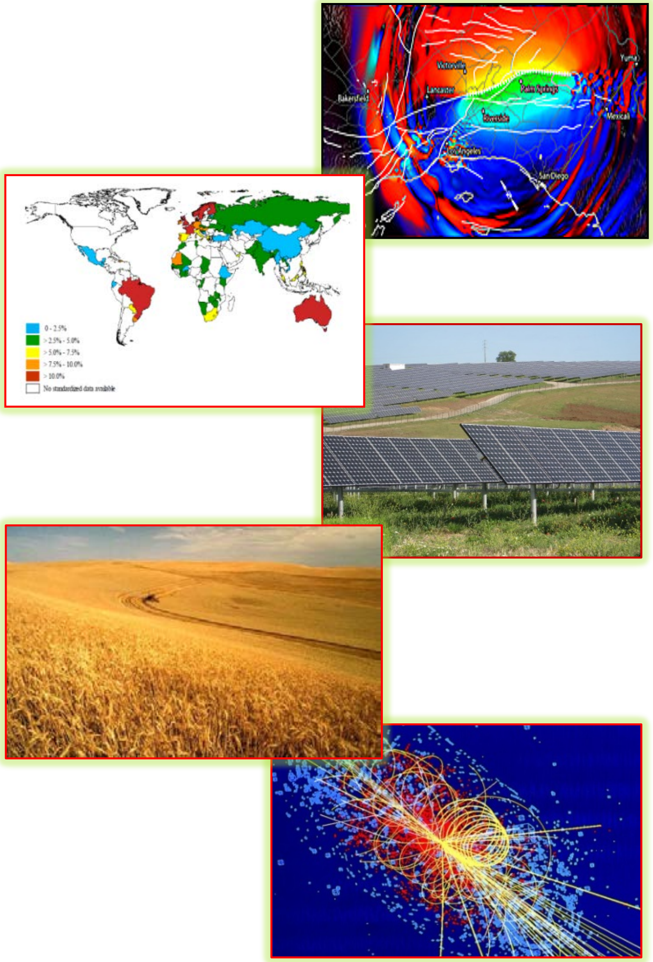
://alliance

Next Steps Towards Sustainable Stewardship

- Identify and evolve an expanding network of repositories for publicly accessible research data

- Create useful metrics for successful stewardship and economic stability that can be used to support the development of effective organizational support

- Create a plan and actionable recommendations for strategic investment



Fran Berman

Last Words: Information Infrastructure is necessary for 21st Century Innovation

- **Value Proposition:**
  - Virtually all fields becoming data driven
  - **Adequate and sustainable data infrastructure is critical to drive innovation and HPC applications**
- **What can we do:**
  - Include data stewardship, management, use, access and preservation as part of project planning, budget and efforts
  - Recognize and publish the data contributions of our work as well as the research contributions

Small steps (things to do on Monday morning)**:**

1. If you don't have one, create a data management plan for your current project for a reasonable fixed term of time
2. Make your data available to the community (as appropriate) by curating it and ingesting it into a publicly accessible repository
3. Cite and publish your data as appropriate when you write about your results

Thank You!

Rensselaer

Fran Berman