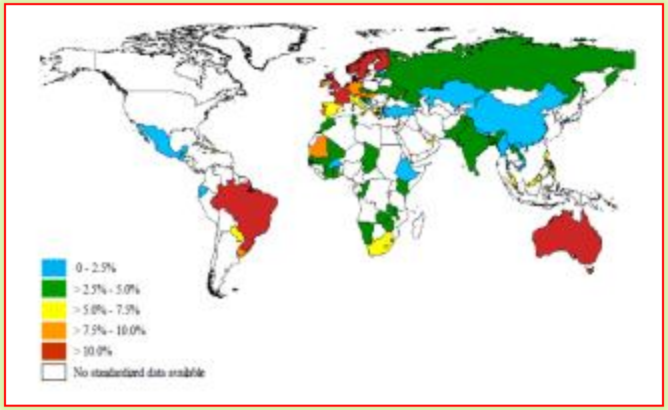
Towards a Sustainable Ecosystem for Data Driven Research and Innovation

Dr. Francine Berman

Chair, Research Data Alliance / US

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

Innovation opportunity: Research data driving solutions to scientific and societal challenges



Who is most at risk to contract asthma?

How can we increase wheat yields?

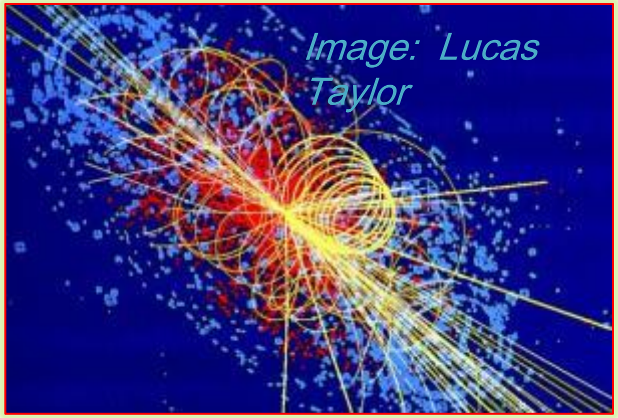


Image: Lucas Taylor

How accurate is the Standard Model of Physics?

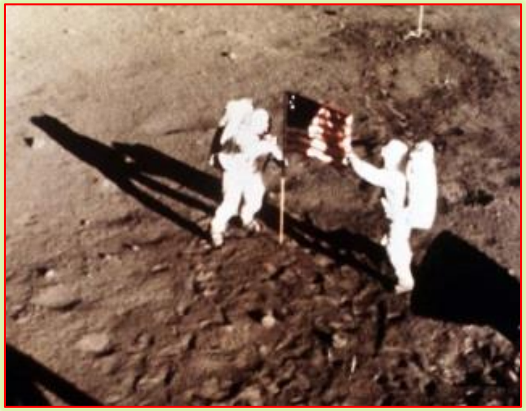
How can we best address energy needs and sustain the environment?

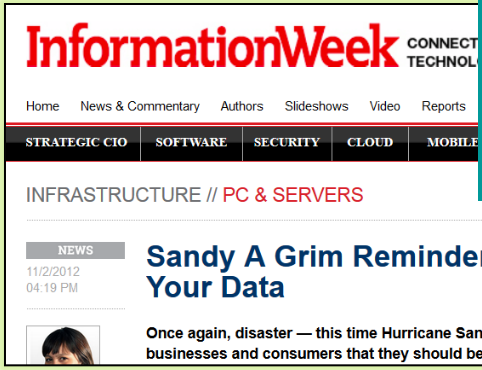
Image: Ceinturion, Wikipedia

Infrastructure reality: Access, use and re-use of data now and in the future presupposes data sustainable stewardship and preservation *today*

- **Stewardship and Preservation critical:**
  "Homeless" data cease to exist

- Sustainable data infrastructure necessary to support
  - Data management plans
  - Public access to research data
  - Use and re-use of data
  - Reproducibility of results





INFRASTRUCTURE // PC & SERVERS

NEWS
11/2/2012
04:19 PM

Sandy A Grim Reminder: Back Up Your Data

Once again, disaster — this time Hurricane Sandy — reminds businesses and consumers that they should be backing up their data.

Rensselaer

NATIONAL SCIENCE FOUNDATION

RDA RESEARCH DATA ALLIANCE

How can we usefully think about sustainability?

Sustainable development: "development that meets the needs of the present without compromising the ability of future generations to meet their own needs."

Our Common Future, U.N. Brundtland Commission

- **Key components**
  - Ecological sustainability
  - Cultural / institutional sustainability
  - Economic sustainability
  - Political sustainability

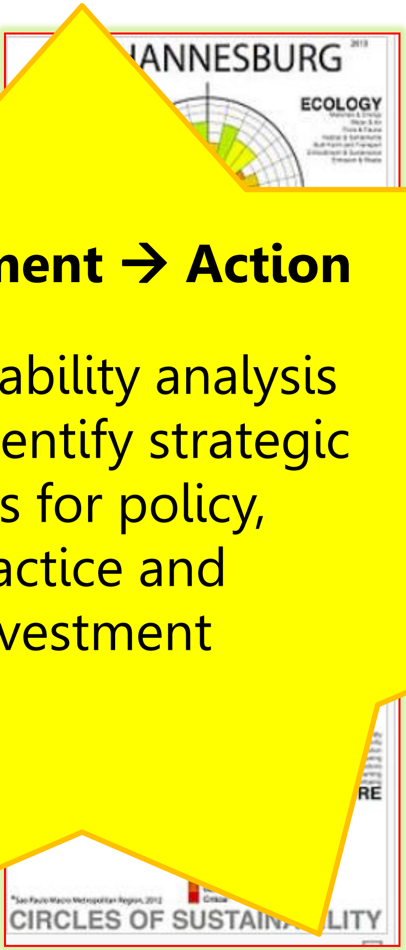
Planet image: NASA; Quote from "Our Common Future" http://www.un-documents.net/our-common-future.pdf

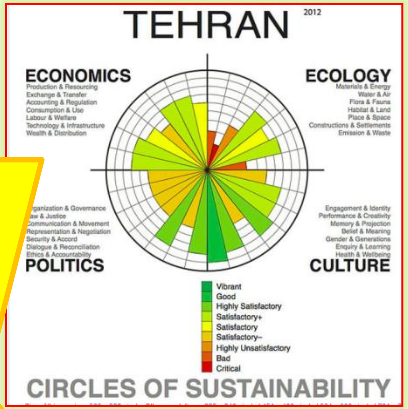
How can we measure sustainability?

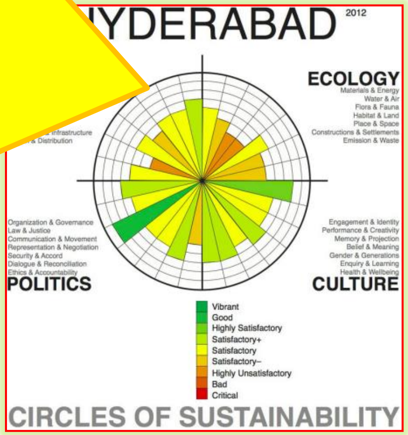
- **Circles of Sustainability** developed to assess and understand sustainability. Used
  - For managing proje towards socially sus outcomes
  - To assess the susta cities and urban
- *Used by global orga including the United Natio Compact Cities Programme, T World Association of Metropol World Vision, and others.*

Assessment → Action

Sustainability analysis helps identify strategic areas for policy, practice and investment



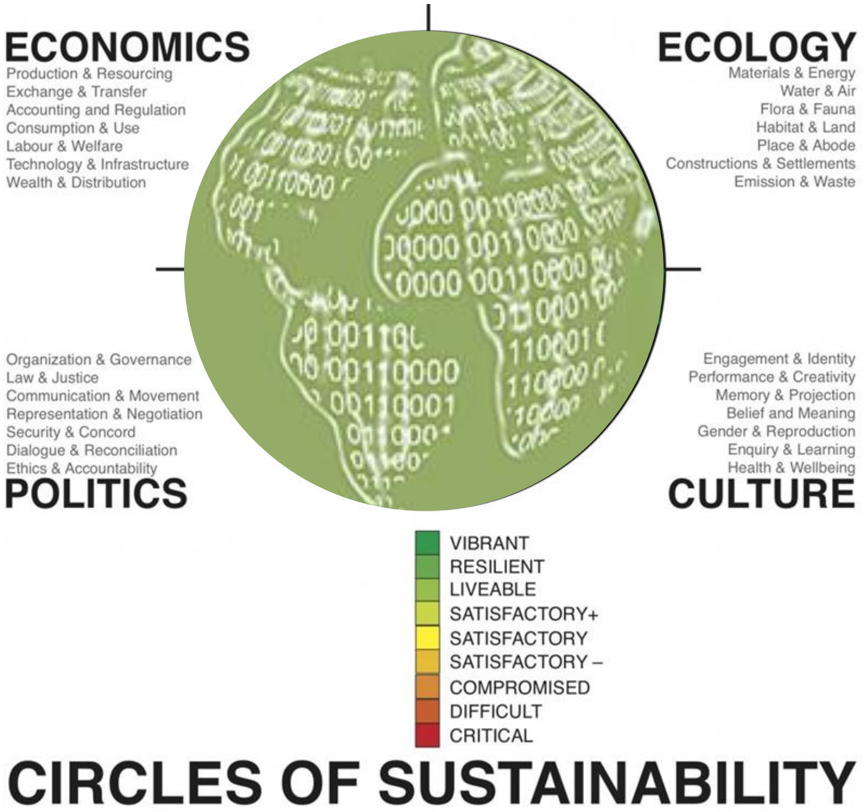




Circles of sustainability provide a useful lens with which to consider sustainability of digital research data

How can we create a viable support model for digital access and preservation?

Why is digital preservation and access such a hard sell?



ECONOMICS

Production & Resourcing
Exchange & Transfer
Accounting and Regulation
Consumption & Use
Labour & Welfare
Technology & Infrastructure
Wealth & Distribution

ECOLOGY

Materials & Energy
Water & Air
Flora & Fauna
Habitat & Land
Place & Abode
Constructions & Settlements
Emission & Waste

POLITICS

Organization & Governance
Law & Justice
Communication & Movement
Representation & Negotiation
Security & Concord
Dialogue & Reconciliation
Ethics & Accountability

CULTURE

Engagement & Identity
Performance & Creativity
Memory & Projection
Belief and Meaning
Gender & Reproduction
Enquiry & Learning
Health & Wellbeing

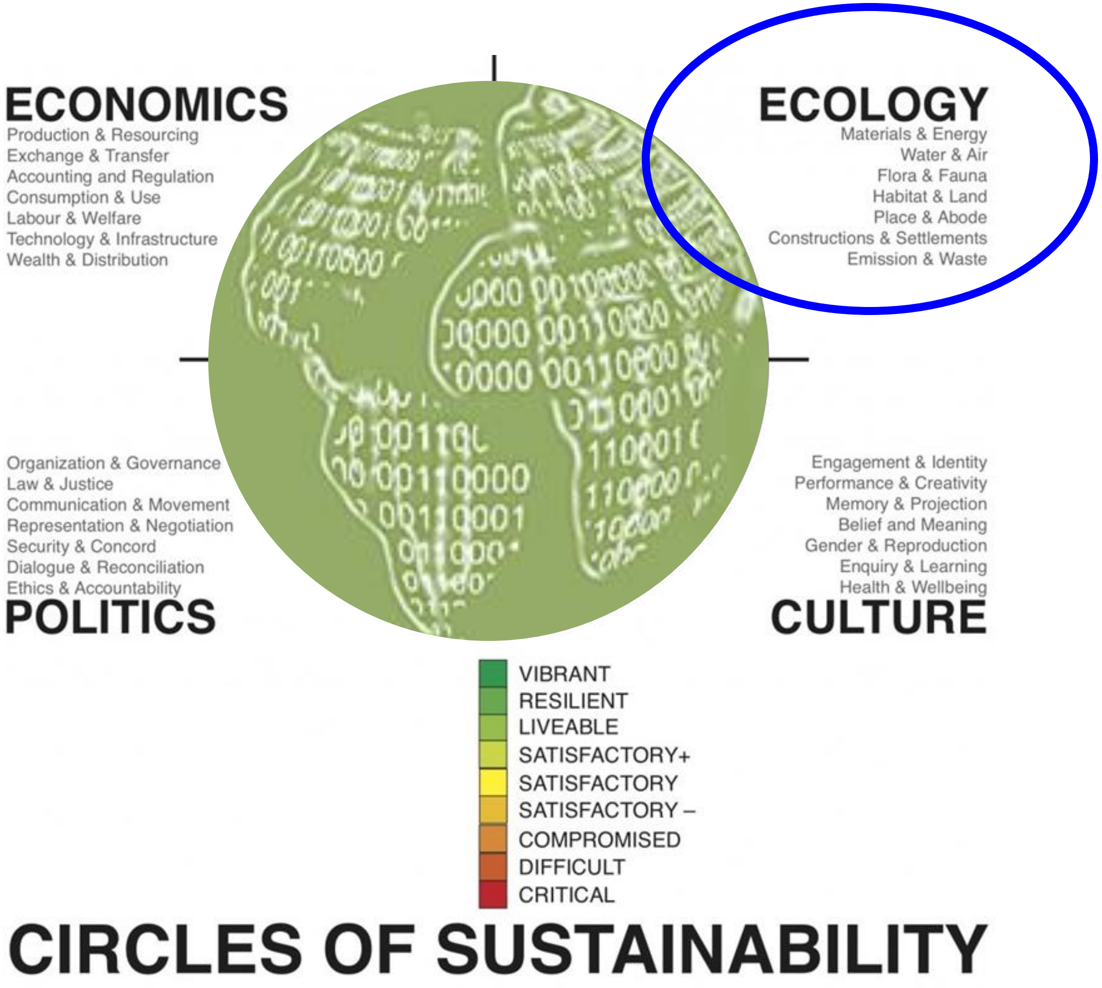
🟩 VIBRANT
🟩 RESILIENT
🟩 LIVEABLE
🟨 SATISFACTORY+
🟨 SATISFACTORY
🟧 SATISFACTORY –
🟧 COMPROMISED
🟥 DIFFICULT
🟥 CRITICAL

CIRCLES OF SUSTAINABILITY

What infrastructure is needed to support digital stewardship and preservation?

How do we maximize the access, sharing and exchange of digital research data?

What infrastructure is needed to support digital stewardship and preservation?



ECONOMICS

Production & Resourcing
Exchange & Transfer
Accounting and Regulation
Consumption & Use
Labour & Welfare
Technology & Infrastructure
Wealth & Distribution

ECOLOGY

Materials & Energy
Water & Air
Flora & Fauna
Habitat & Land
Place & Abode
Constructions & Settlements
Emission & Waste

Organization & Governance
Law & Justice
Communication & Movement
Representation & Negotiation
Security & Concord
Dialogue & Reconciliation
Ethics & Accountability

POLITICS

Engagement & Identity
Performance & Creativity
Memory & Projection
Belief and Meaning
Gender & Reproduction
Enquiry & Learning
Health & Wellbeing

CULTURE

🟩 VIBRANT
🟩 RESILIENT
🟩 LIVEABLE
🟨 SATISFACTORY+
🟨 SATISFACTORY
🟨 SATISFACTORY –
🟧 COMPROMISED
🟧 DIFFICULT
🟥 CRITICAL

CIRCLES OF SUSTAINABILITY

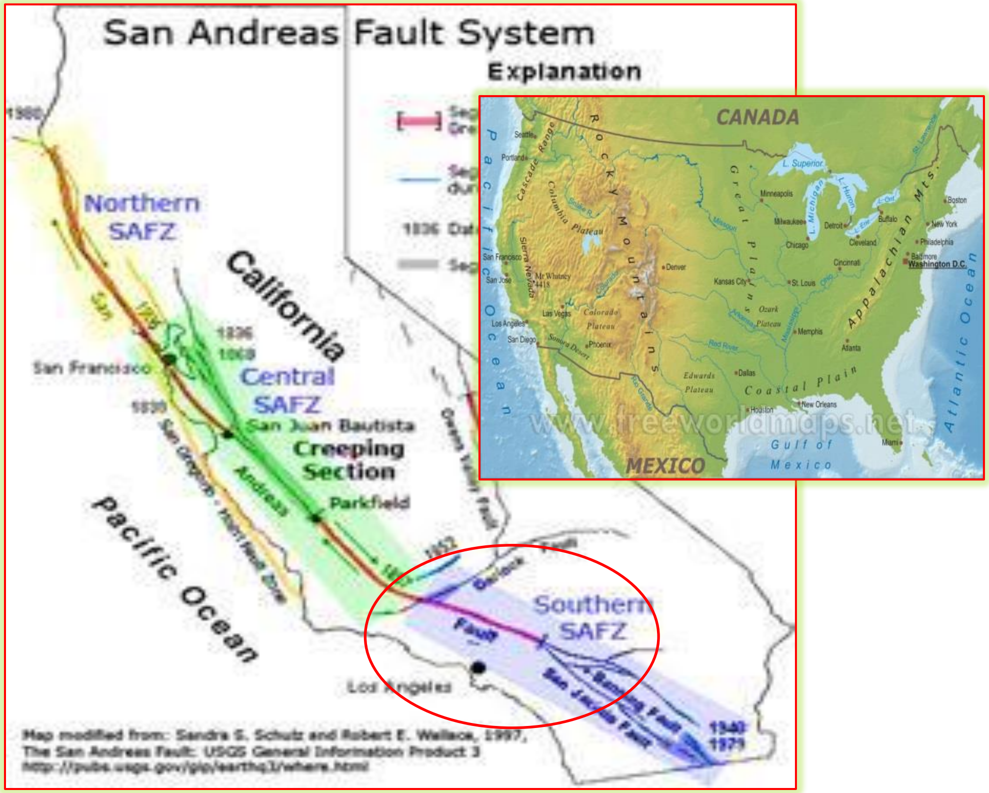
Data-driven Geoscience: How can we respond to large-scale earthquakes?

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



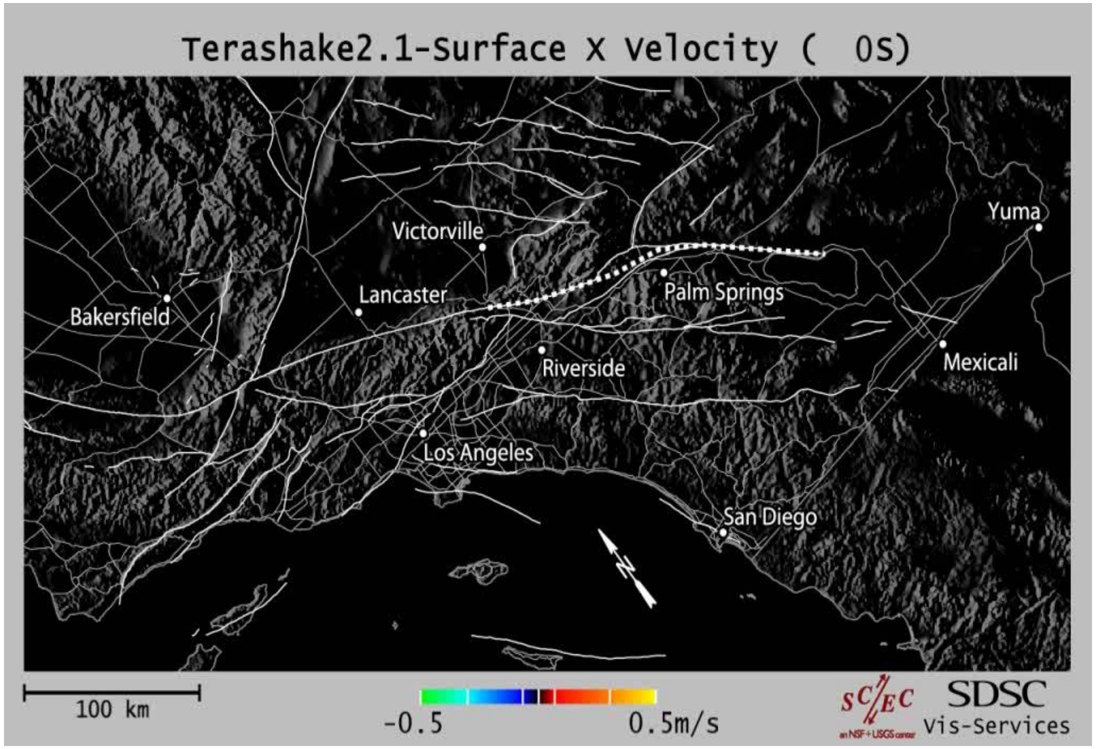
Map modified from: Sandra S. Schulz and Robert E. Wallace, 1997, The San Andreas Fault: USGS General Information Product 3 http://pubs.usgs.gov/gip/earthq3/where.html

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

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

Terashake data infrastructure*

- **Data Management**
  - 10 Terabytes moved per day during execution over 5 days
  - Derived data products registered into SCEC digital library (total SCEC library has 168 TB)
- **Data Post-processing:**
  - *Movies* of seismic wave propagation
  - Seismogram formatting for interactive on-line analysis
  - *Derived data:*
    - Velocity magnitude
    - Displacement vector field
    - Cumulative peak maps
    - Statistics used in visualizations

* circa ~2007

TeraShake Resources

Computers and Systems

- 80,000 hours on IBM Power 4 (DataStar)
- 256 GB memory p690 used for testing, p655s used for production run, TeraGrid used for porting
- 30 TB Global Parallel file GPFS
- Run-time 100 MB/s data transfer from GPFS to SAM-QFS
- 27,000 hours post-processing for high resolution rendering

People

- 20+ people for IT support
- 20+ people in domain research

Storage

- SAM-QFS archival storage
- HPSS backup
- Storage Resource Broker collection with 1,000,000 files

Technical infrastructure critical part of the ecosystem for data-driven innovation

Data access via portals, science gateways, etc.

Database and data collection systems

Data services to support use and re-use

Data analysis algorithms, data-driven models and simulations

Data visualization tools

Semantic frameworks

Data management systems

Data storage

Social, organizational, and human infrastructure for data-driven results equally important

Policy

Systems Interoperability

Common Standards

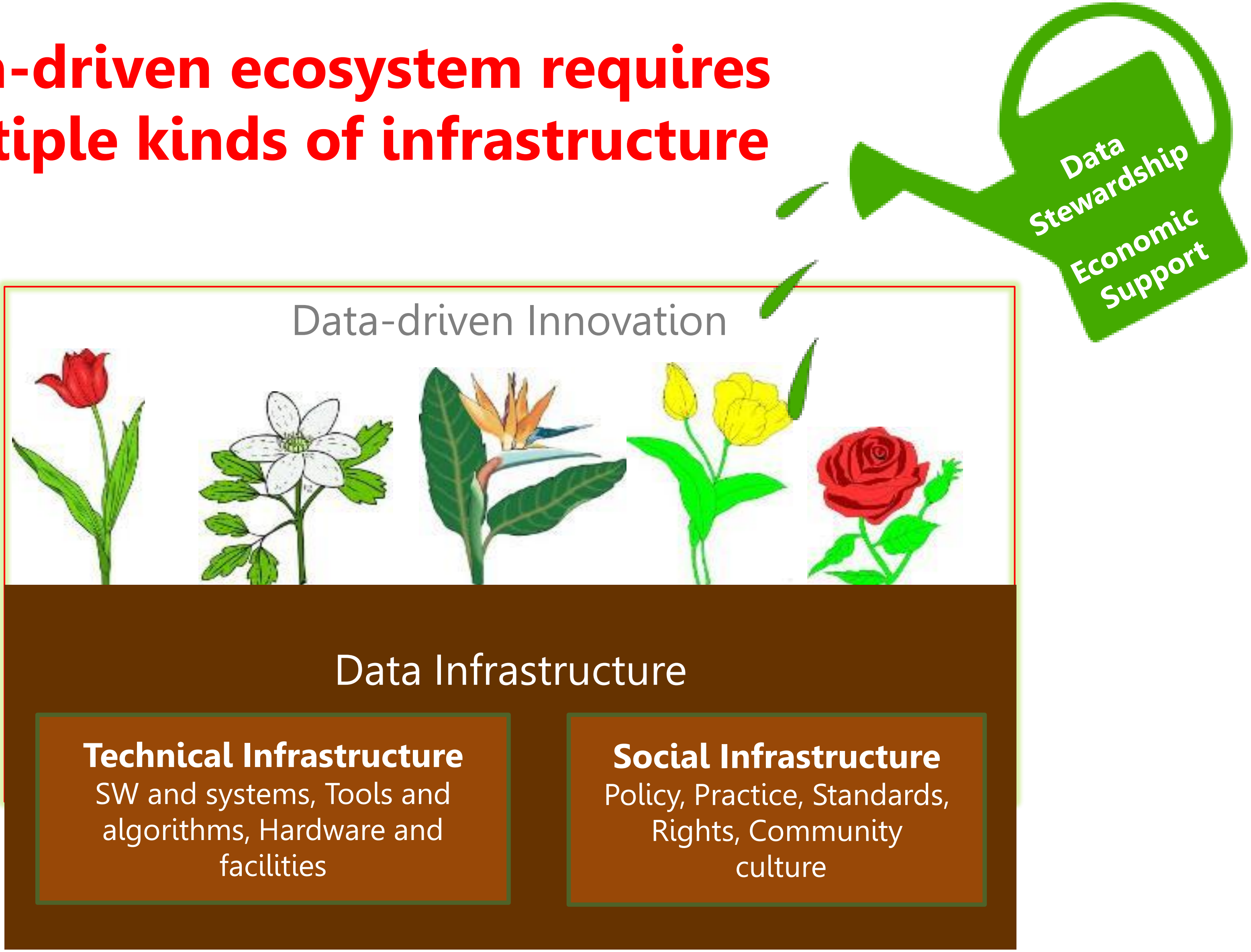
Sustainable Economics

Community Practice

Traffic Image: Mike Gonzalez

Workforce and training

Data-driven ecosystem requires multiple kinds of infrastructure



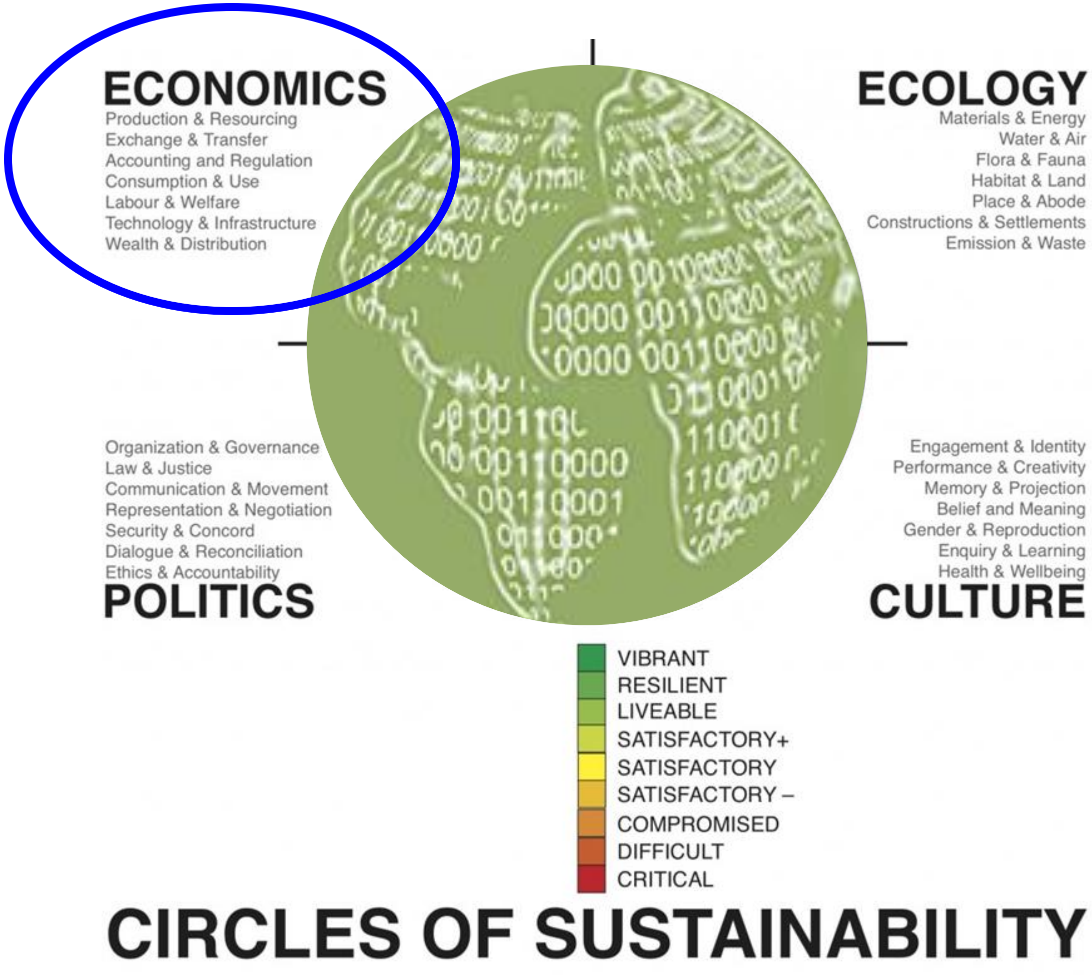
Data-driven Innovation

Data Infrastructure

Technical Infrastructure
SW and systems, Tools and algorithms, Hardware and facilities

Social Infrastructure
Policy, Practice, Standards, Rights, Community culture

How can we create a viable support model for digital access and preservation?



ECONOMICS

Production & Resourcing
Exchange & Transfer
Accounting and Regulation
Consumption & Use
Labour & Welfare
Technology & Infrastructure
Wealth & Distribution

ECOLOGY

Materials & Energy
Water & Air
Flora & Fauna
Habitat & Land
Place & Abode
Constructions & Settlements
Emission & Waste

Organization & Governance
Law & Justice
Communication & Movement
Representation & Negotiation
Security & Concord
Dialogue & Reconciliation
Ethics & Accountability

POLITICS

Engagement & Identity
Performance & Creativity
Memory & Projection
Belief and Meaning
Gender & Reproduction
Enquiry & Learning
Health & Wellbeing

CULTURE

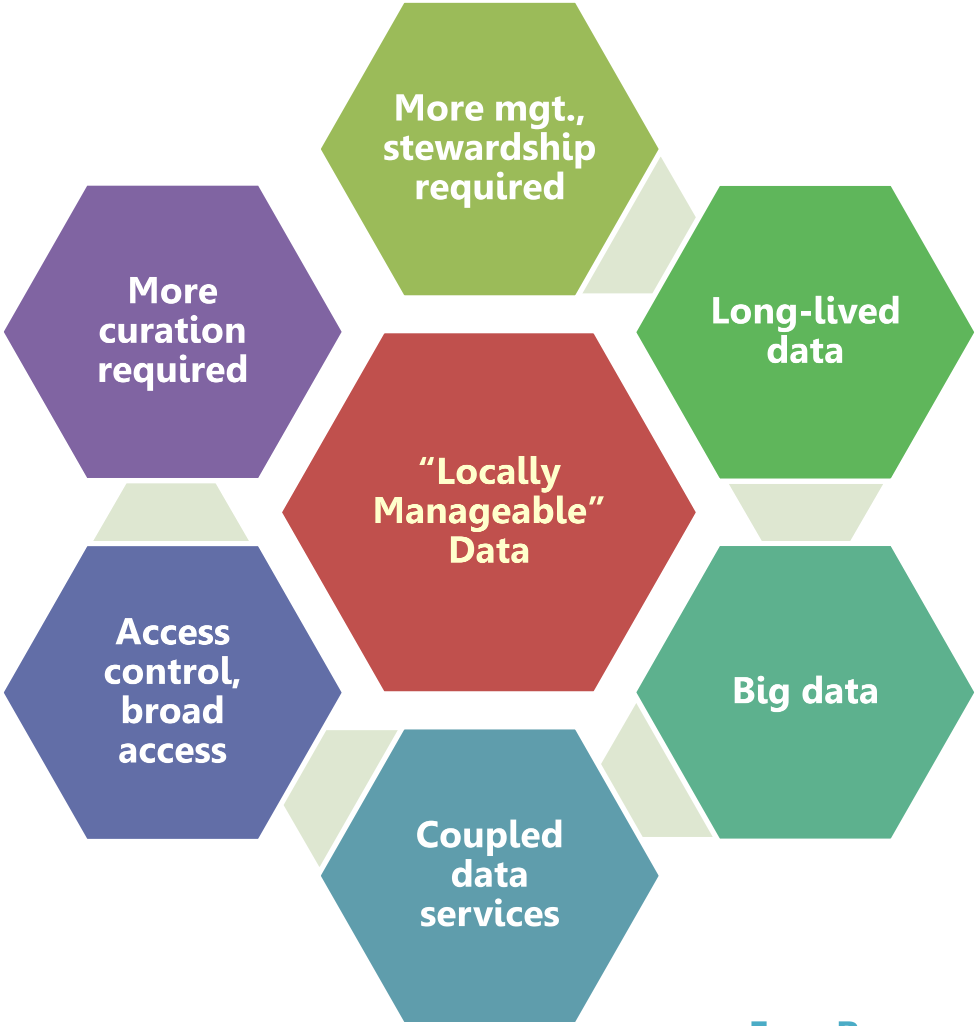
🟩 VIBRANT
🟩 RESILIENT
🟩 LIVEABLE
🟨 SATISFACTORY+
🟨 SATISFACTORY
🟧 SATISFACTORY –
🟧 COMPROMISED
🟥 DIFFICULT
🟥 CRITICAL

CIRCLES OF SUSTAINABILITY

Data economics: Responsible data stewardship requires a viable business model for sustaining its underlying infrastructure

Data infrastructure costs increase with usage, stewardship and access requirements, perceived value

Greater costs at the extremes (including "big" data) …



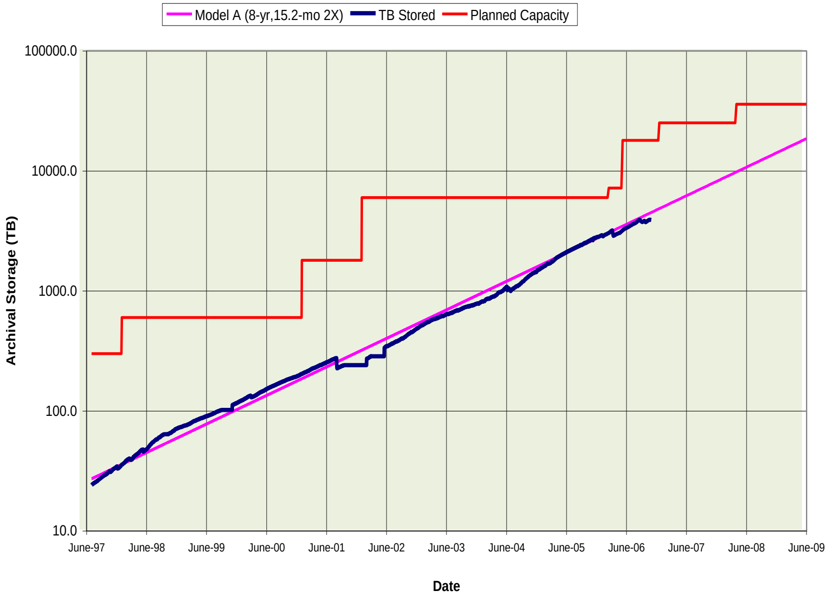
More mgt., stewardship required
More curation required
Long-lived data
"Locally Manageable" Data
Access control, broad access
Big data
Coupled data services

It’s not just about the cost of storage

Data Infrastructure 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, policy, etc. …

Resources and Resource Refresh



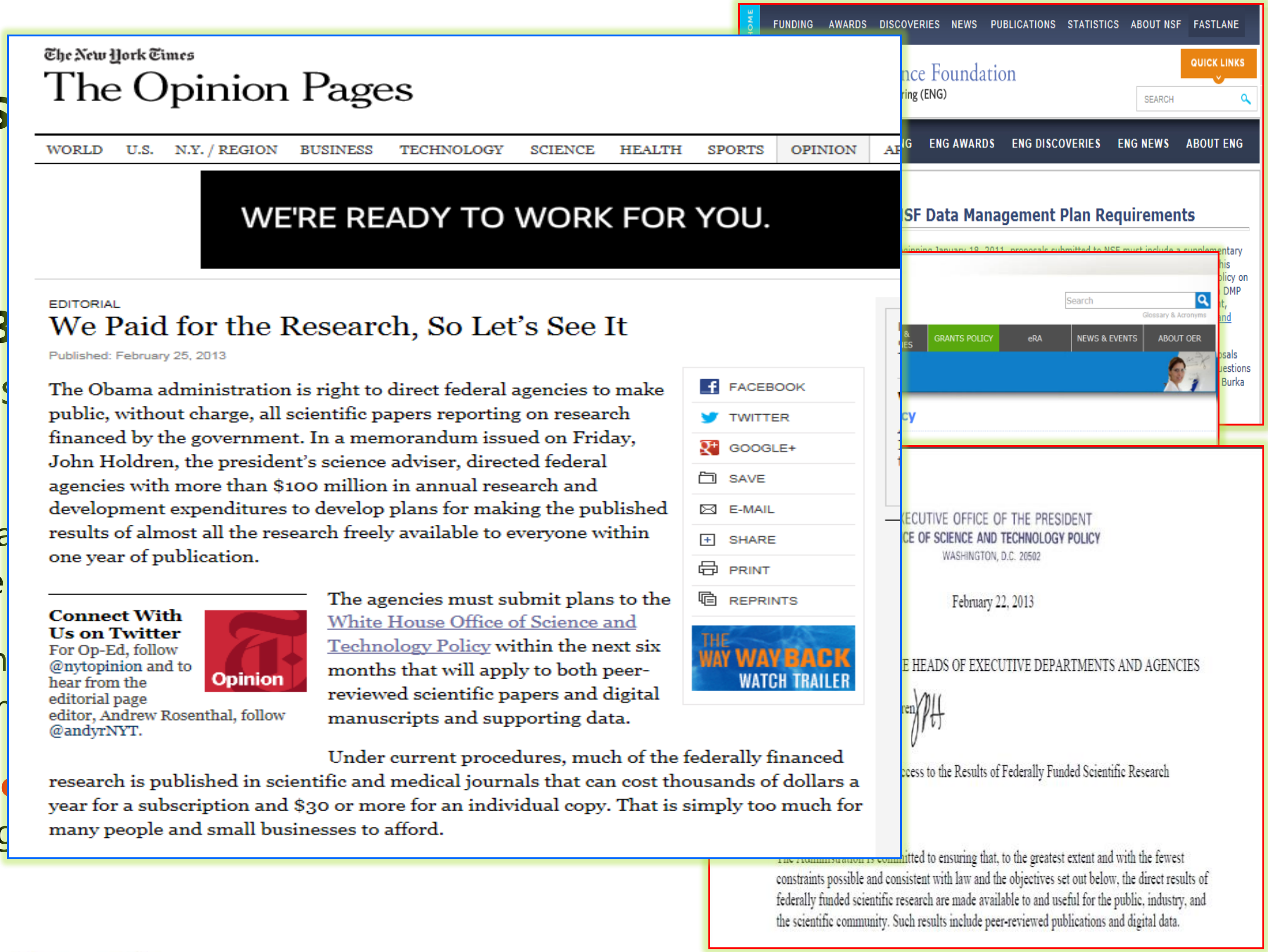
SDSC Data Storage Growth ‘97-’09

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

Information courtesy of Richard Moore, SDSC

Increased requirements for access mean increased need for data infrastructure

- **Increasing U.S** **requirements** research data
- **February 2013** for public acces publications
  - Strategy for ca public-private
  - Strategy for in access, dissem
  - **No new mone** the existing ag



The New York Times

The Opinion Pages

EDITORIAL

We Paid for the Research, So Let's See It

Published: February 25, 2013

The Obama administration is right to direct federal agencies to make public, without charge, all scientific papers reporting on research financed by the government. In a memorandum issued on Friday, John Holdren, the president's science adviser, directed federal agencies with more than $100 million in annual research and development expenditures to develop plans for making the published results of almost all the research freely available to everyone within one year of publication.

Connect With Us on Twitter
For Op-Ed, follow @nytopinion and to hear from the editorial page editor, Andrew Rosenthal, follow @andyrNYT.

The agencies must submit plans to the White House Office of Science and Technology Policy within the next six months that will apply to both peer-reviewed scientific papers and digital manuscripts and supporting data.

Under current procedures, much of the federally financed research is published in scientific and medical journals that can cost thousands of dollars a year for a subscription and $30 or more for an individual copy. That is simply too much for many people and small businesses to afford.

SF Data Management Plan Requirements

ginning January 18, 2011, proposals submitted to NSF must include a supplementary his licy on DMP t, nd osals uestions Burka

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

February 22, 2013

E HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

ren JPH

ccess to the Results of Federally Funded Scientific Research

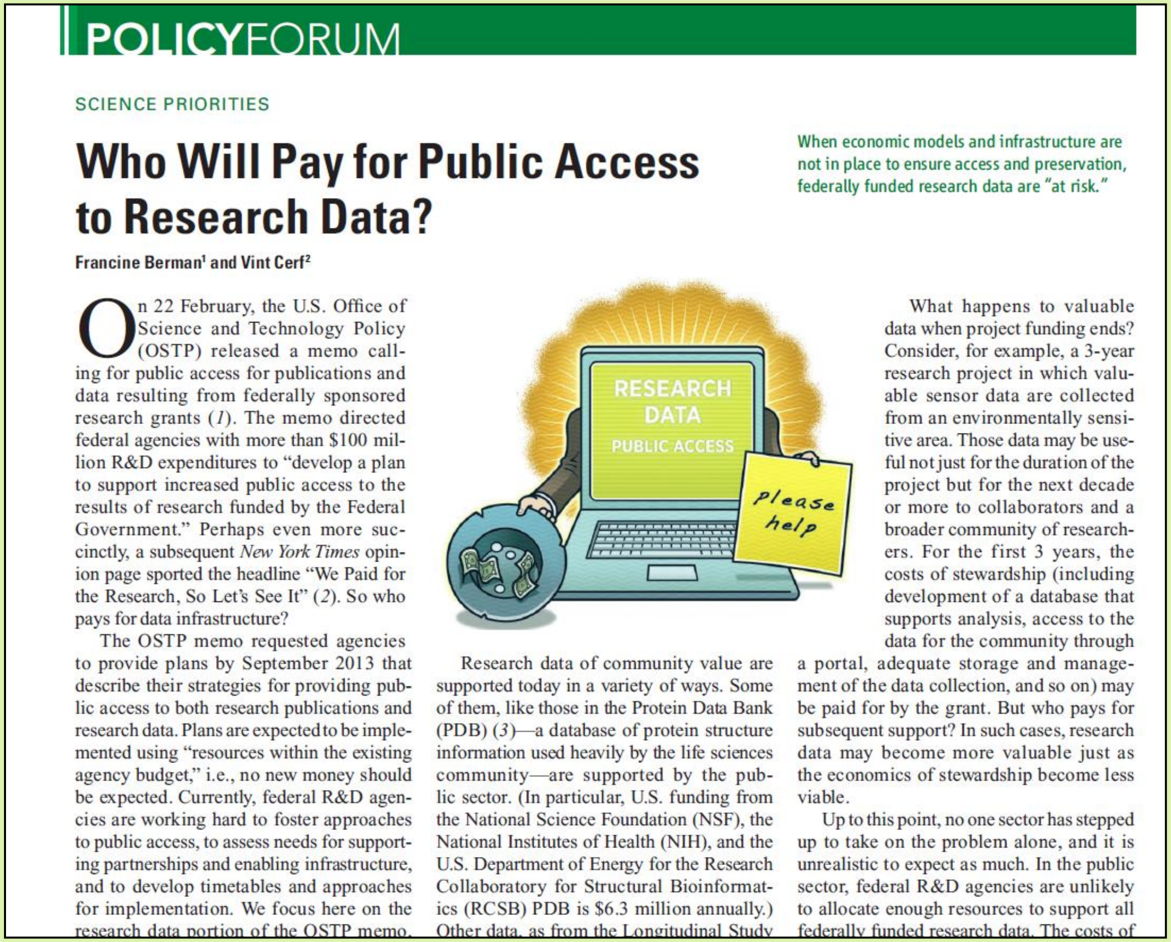
The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific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

Rensselaer

NATIONAL SCIENCE FOUNDATION

RDA RESEARCH DATA ALLIANCE

Economics of public access:
Who pays the data bill?


RESEARCH DATA
PUBLIC ACCESS
please help

POLICYFORUM

SCIENCE PRIORITIES

Who Will Pay for Public Access to Research Data?

Francine Berman[1] and Vint Cerf[2]

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

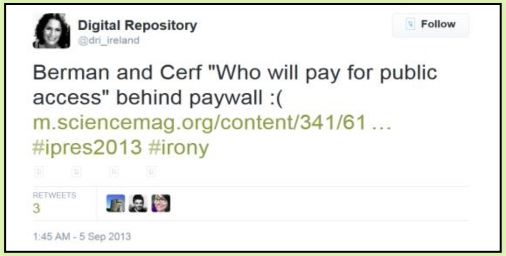
On 22 February, the U.S. Office of Science and Technology Policy (OSTP) released a memo calling for public access for publications and data resulting from federally sponsored research grants (*1*). The memo directed federal agencies with more than $100 million R&D expenditures to "develop a plan to support 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 has stepped up to take on the problem alone, and it is unrealistic to expect as much. In the public sector, federal R&D agencies are unlikely to allocate enough resources to support all federally funded research data. The costs of

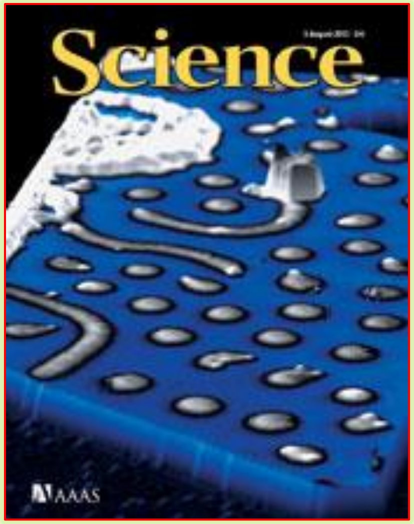


Digital Repository
@dri_ireland

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

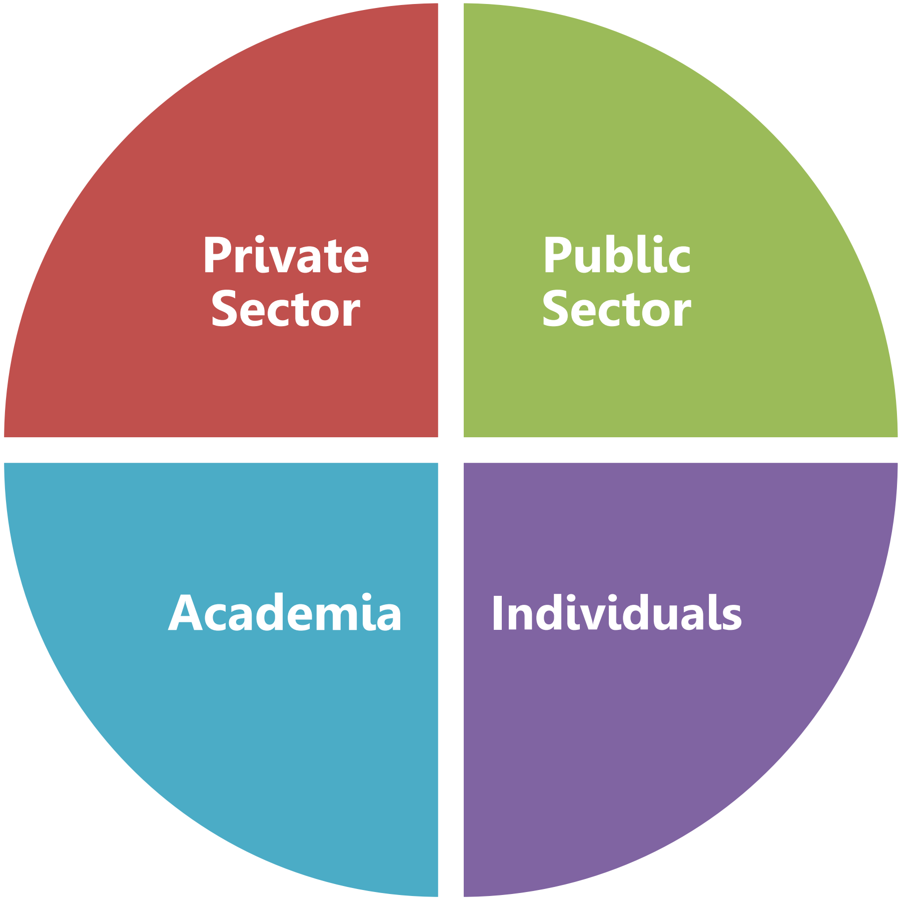
RETWEETS
3

1:45 AM - 5 Sep 2013



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

Op-ed recommendations: Cultivate / coordinate preservation and stewardship options in every sector



Private Sector

Public Sector

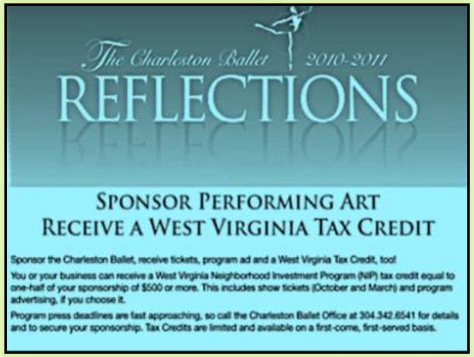
Academia

Individuals

Op-ed recommendations: Cultivate / coordinate preservation and stewardship options in every sector

Private Sector

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



The Charleston Ballet 2010-2011

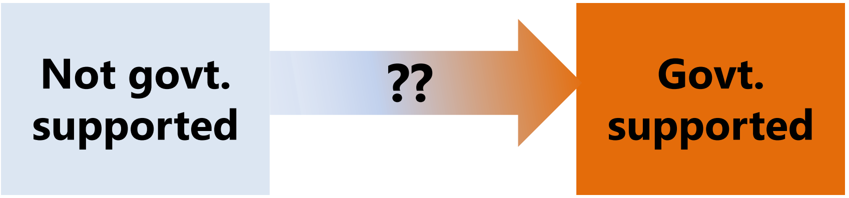
REFLECTIONS

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.



Not govt. supported

??

Govt. supported

Public Sector

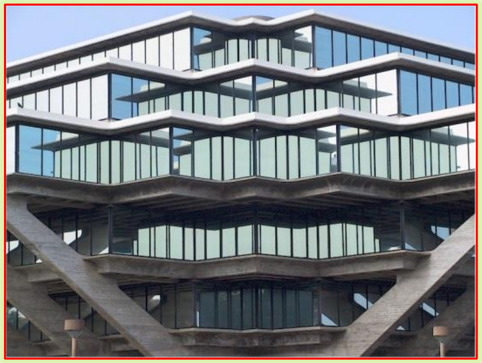
- Clarify public sector stewardship commitments: articulate what data will / won't be supported

Charleston Ballet blog: http://allianceblog.org/tag/charleston-ballet/ ; corporate and collection logos

Op-ed recommendations: Cultivate / coordinate preservation and stewardship options in every sector

Academic Sector

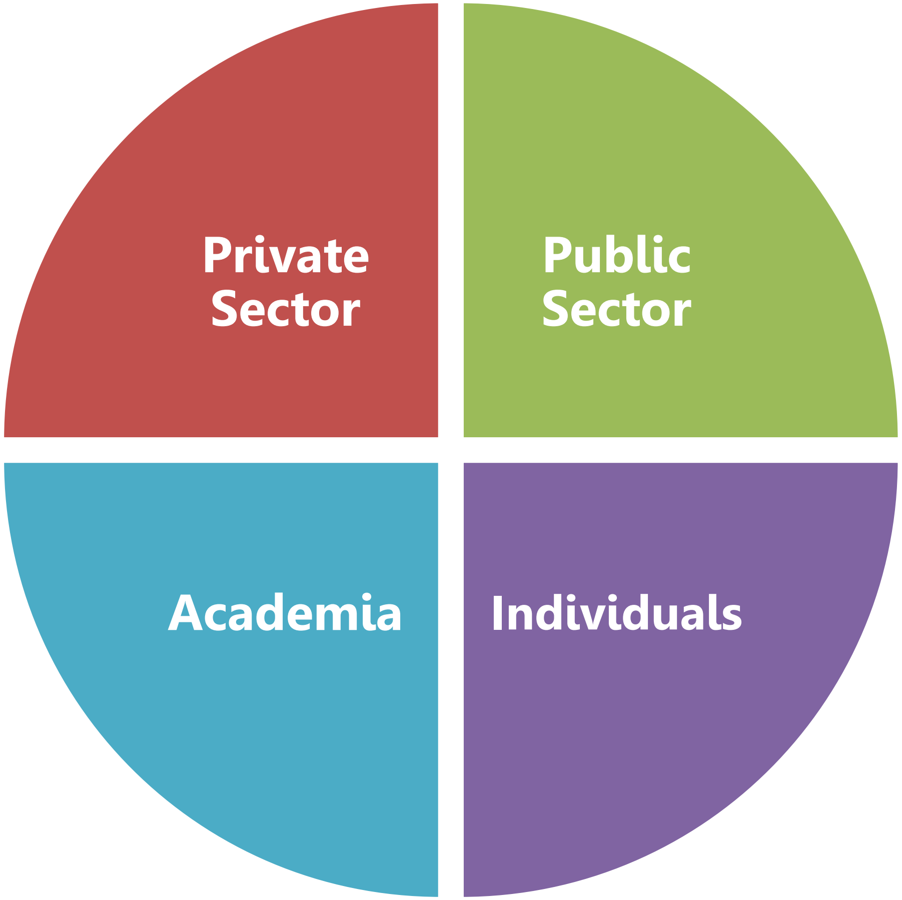
- Create sustainable university library and repository stewardship solutions



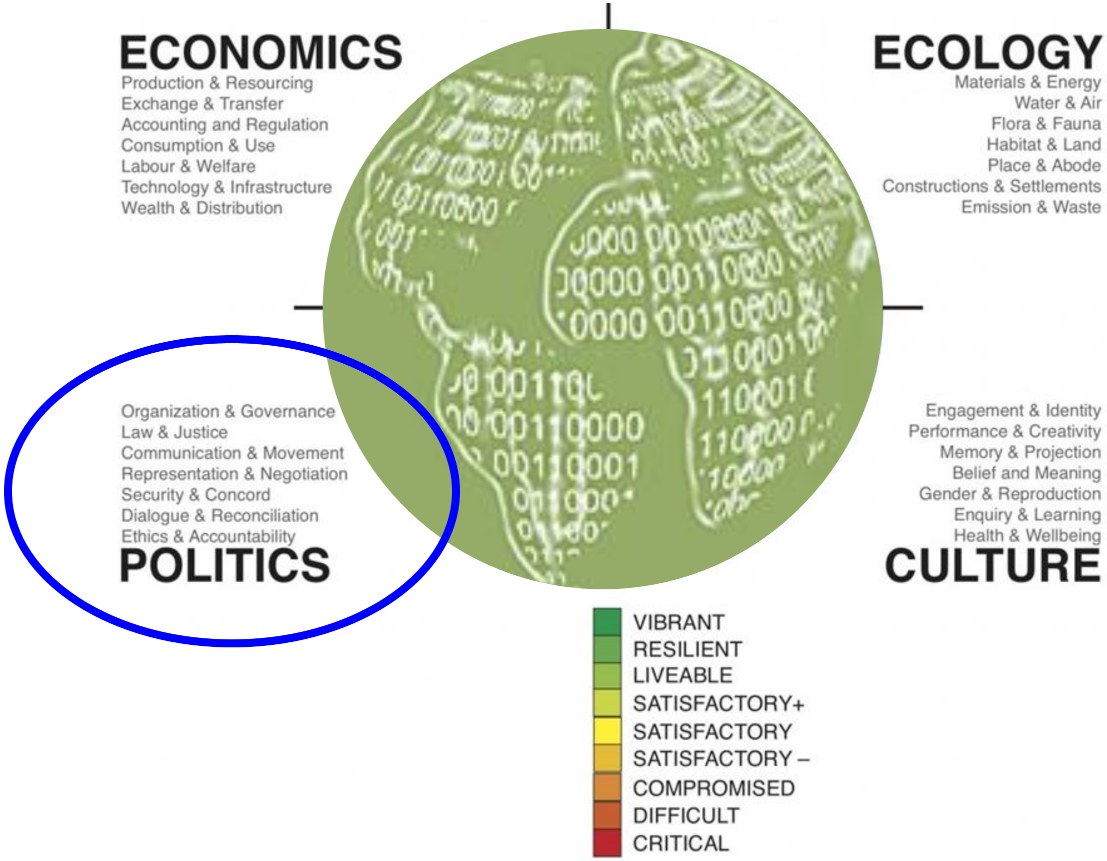
Individuals

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

No magic economic bullet. Coordination between approaches can provide even more robust options for stewardship



Why is digital preservation and access such a hard sell?



ECONOMICS
Production & Resourcing
Exchange & Transfer
Accounting and Regulation
Consumption & Use
Labour & Welfare
Technology & Infrastructure
Wealth & Distribution

ECOLOGY
Materials & Energy
Water & Air
Flora & Fauna
Habitat & Land
Place & Abode
Constructions & Settlements
Emission & Waste

Organization & Governance
Law & Justice
Communication & Movement
Representation & Negotiation
Security & Concord
Dialogue & Reconciliation
Ethics & Accountability
POLITICS

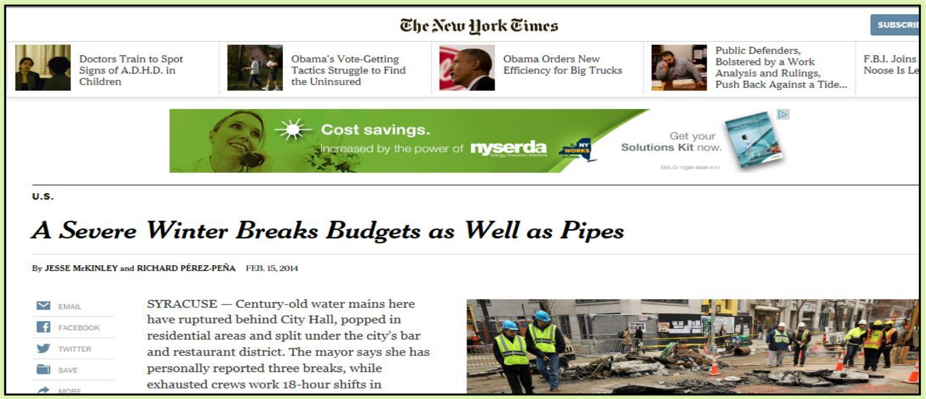
Engagement & Identity
Performance & Creativity
Memory & Projection
Belief and Meaning
Gender & Reproduction
Enquiry & Learning
Health & Wellbeing
CULTURE

🟩 VIBRANT
🟩 RESILIENT
🟩 LIVEABLE
🟨 SATISFACTORY+
🟨 SATISFACTORY
🟧 SATISFACTORY –
🟧 COMPROMISED
🟥 DIFFICULT
🟥 CRITICAL

CIRCLES OF SUSTAINABILITY

Academic and public sector infrastructure challenges

	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
What is the funding model?	Fixed-term funding	Continuous long-term support
Who is responsible?	Various govt. R&D agencies, NGOs, etc.	No-one's major priority


SUBSCRI
Cost savings.
Increased by the power of nyserda
NY WORKS
Get your Solutions Kit now.
EMAIL
FACEBOOK
TWITTER
SAVE
MORE

The New York Times

Doctors Train to Spot Signs of A.D.H.D. in Children

Obama's Vote-Getting Tactics Struggle to Find the Uninsured

Obama Orders New Efficiency for Big Trucks

Public Defenders, Bolstered by a Work Analysis and Rulings, Push Back Against a Tide...

F.B.I. Joins Noose Is Le

U.S.

A Severe Winter Breaks Budgets as Well as Pipes

By JESSE McKINLEY and RICHARD PÉREZ-PEÑA FEB. 15, 2014

SYRACUSE — Century-old water mains here have ruptured behind City Hall, popped in residential areas and split under the city's bar and restaurant district. The mayor says she has personally reported three breaks, while exhausted crews work 18-hour shifts in

Stephanie A. Miner, the Syracuse mayor, said [infrastructure is] too often overlooked when politicians want to spend money on economic development. ***"You don't cut ribbons for new water mains, but that's really what matters."***

NY Times, Feburary 15, 2014

Systemic challenges to sustainable stewardship

(from the Blue Ribbon Task Force Interim Report [at brtf.sdsc.edu])

- **Poor alignment between stakeholders** in the digital preservation and access world and their roles, responsibilities and support models
- There is a **lack of institutional, enterprise, and/or community incentives** to support the collaboration needed to enforce sustainable economic models
- **Complacency that current practices are "good enough"** and / or the **problem is not urgent.** Both "carrots" (in the form of recognition that access to information is an investment in current and future success) and "sticks" (in the form of penalties for non-compliance, accounting of explicit opportunity costs, or costs of lost information) are needed
- Fear that digital access and **preservation is too big to take on**

Reports are not enough, but good reports can provide compelling evidence needed by stakeholders for action

THE DIGITAL DILEMMA

STRATEGIC ISSUES IN ARCHIVING AND ACCESSING DIGITAL MOTION PICTURE MATERIALS

Science and Technology Council

DIGITAL PRESERVATION POLICIES STUDY

Part 1: Final Report October 2008

Neil Beagrie, Najla Semple, Peter Williams, and Richard Wright

Prepared by:
Charles Beagrie Limited
www.beagrie.com
A study funded by
JISC
Copyright HEFCE 2008
The authors have asserted their moral rights in this work

Sustaining the Digital Investment

Issues and Challenges of Economically Sustainable Digital Preservation

Interim Report of the
Blue Ribbon Task Force on
Sustainable Digital Preservation and Access

An IDC White Paper - sponsored by EMC

The Diverse and Exploding Digital Universe

An Updated Forecast of Worldwide Information Growth Through 2011

March 2008

John F. Gantz, Project Director
Christopher Chute
Alex Manfrediz
Stephen Minton
David Reinsel
ng Schlichting
Anna Toncheva

HARNESSING THE POWER of DIGITAL DATA for SCIENCE AND SOCIETY

Report of the Interagency Working Group on Digital Data
to the Committee on Science of the National Science and Technology Council
January 2009

Investing in Curation

A Shared Path to Sustainability
(Draft)

4C

McKinsey&Company

McKinsey Global Institute

Big data: The next frontier for innovation, competition, and productivity

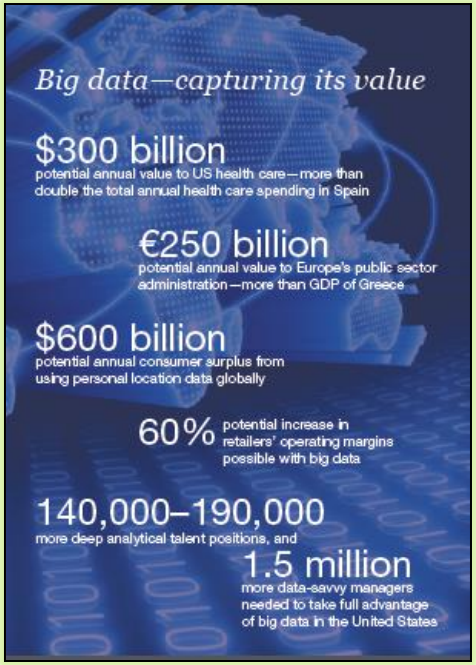
Rensselaer

NATIONAL SCIENCE FOUNDATION

RDA
RESEARCH DATA ALLIANCE

Making the case: Quantifying / qualifying advancement and fear

- **Political capital critical for prioritization and investment in digital stewardship and preservation**
- *Arguments that influence stakeholders and their enablers:*
  - Better economic growth / more jobs
  - Greater leadership / accelerated innovation
  - Increased reputation / competitive advantage
  - Fear of disaster / loss of reputation



Big data—capturing its value

$300 billion
potential annual value to US health care—more than double the total annual health care spending in Spain

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

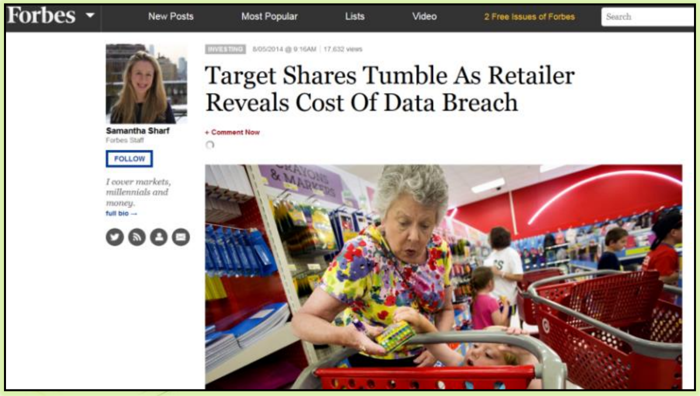
$600 billion
potential annual consumer surplus from using personal location data globally

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

Big Data: The next frontier for innovation, competition and productivity
McKinsey Global Institute, 2011



INVESTING 8/05/2014 @ 9:16AM | 17,632 views

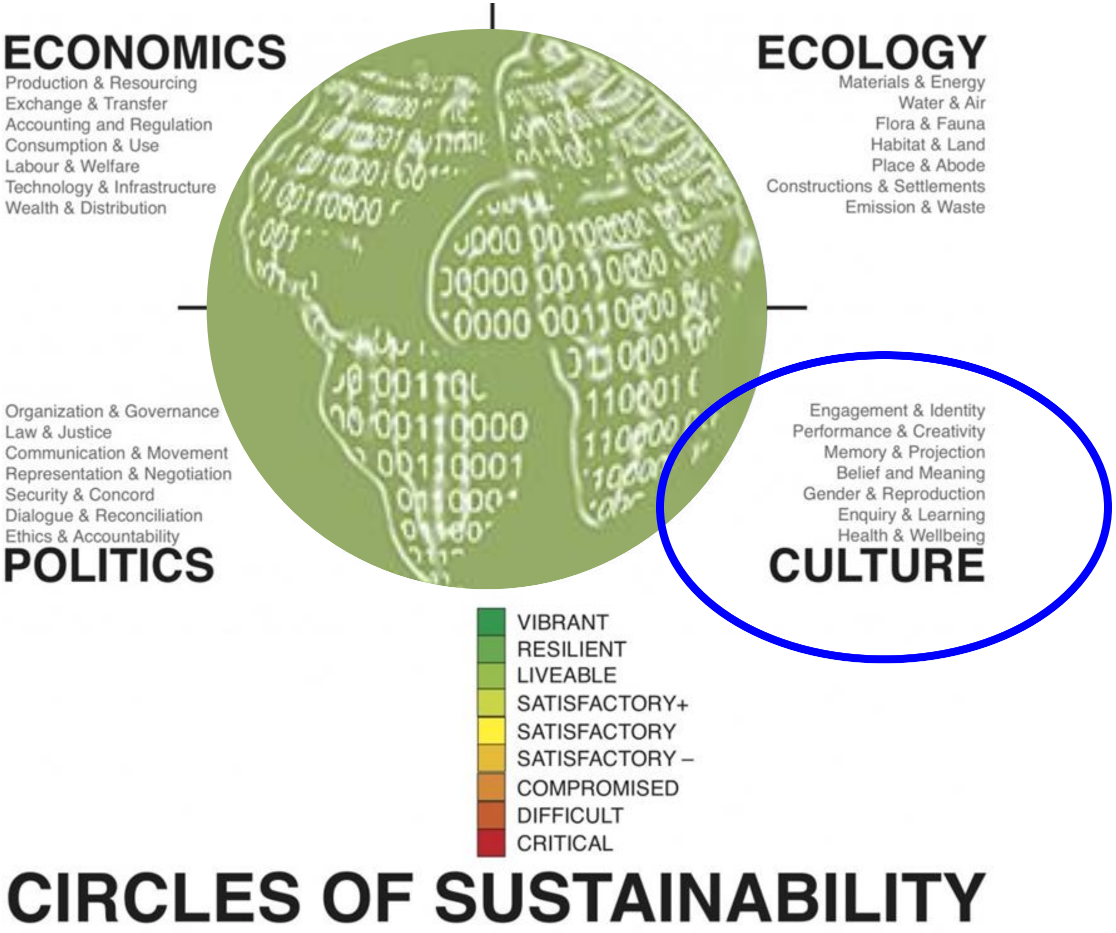
Target Shares Tumble As Retailer Reveals Cost Of Data Breach

Samantha Sharf
Forbes Staff

I cover markets, millennials and money.
full bio →

+ Comment Now

How do we maximize the access, sharing and exchange of digital research data?



ECONOMICS

Production & Resourcing
Exchange & Transfer
Accounting and Regulation
Consumption & Use
Labour & Welfare
Technology & Infrastructure
Wealth & Distribution

ECOLOGY

Materials & Energy
Water & Air
Flora & Fauna
Habitat & Land
Place & Abode
Constructions & Settlements
Emission & Waste

Organization & Governance
Law & Justice
Communication & Movement
Representation & Negotiation
Security & Concord
Dialogue & Reconciliation
Ethics & Accountability

POLITICS

Engagement & Identity
Performance & Creativity
Memory & Projection
Belief and Meaning
Gender & Reproduction
Enquiry & Learning
Health & Wellbeing

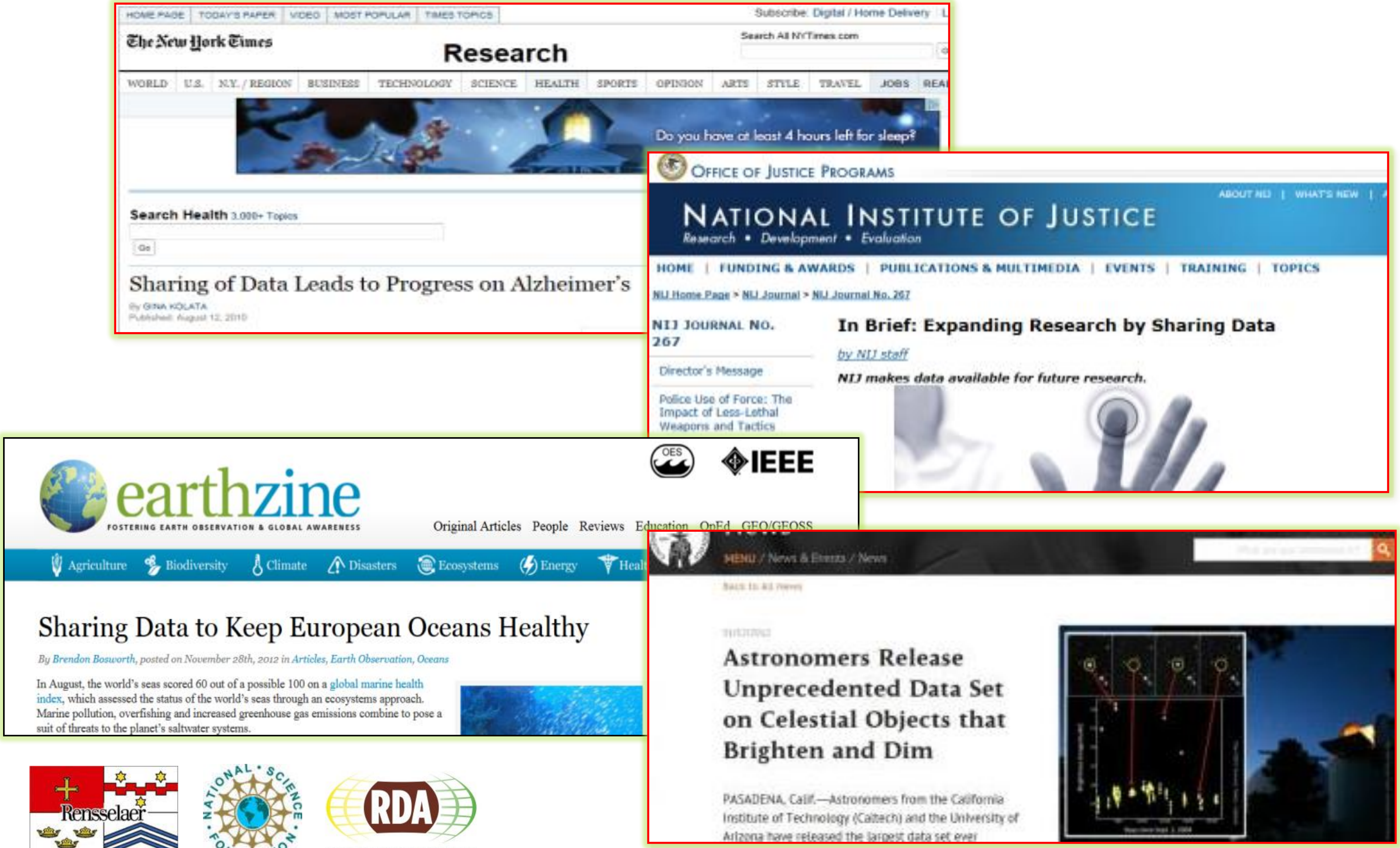
CULTURE

VIBRANT
RESILIENT
LIVEABLE
SATISFACTORY+
SATISFACTORY
SATISFACTORY –
COMPROMISED
DIFFICULT
CRITICAL

CIRCLES OF SUSTAINABILITY

RESEARCH DATA ALLIANCE

Data-sharing driving innovation across research cultures



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 | REAL

Do you have at least 4 hours left for sleep?

Search Health 3,000+ Topics

Go

Sharing of Data Leads to Progress on Alzheimer's

By GINA KOLATA
Published: August 12, 2010

OFFICE OF JUSTICE PROGRAMS

NATIONAL INSTITUTE OF JUSTICE
Research • Development • Evaluation

ABOUT NIJ | WHAT'S NEW |

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

In Brief: Expanding Research by Sharing Data

by NIJ staff

NIJ makes data available for future research.

earthzine
FOSTERING EARTH OBSERVATION & GLOBAL AWARENESS

OES IEEE

Original Articles | People | Reviews | Education | OpEd | GEO/GEOSS

Agriculture | Biodiversity | Climate | Disasters | Ecosystems | Energy | Heal

Sharing Data to Keep European Oceans Healthy

By Brendon Bosworth, posted on November 28th, 2012 in Articles, Earth Observation, Oceans

In August, the world's seas scored 60 out of a possible 100 on a global marine health index, which assessed the status of the world's seas through an ecosystems approach. Marine pollution, overfishing and increased greenhouse gas emissions combine to pose a suit of threats to the planet's saltwater systems.

MENU / News & Events / News

Back to All News

[illegible]

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

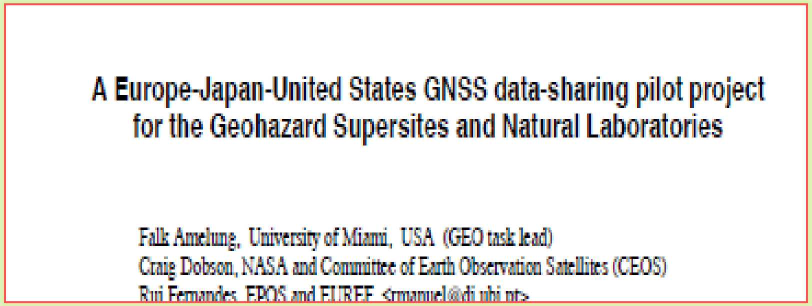
PASADENA, Calif.—Astronomers from the California Institute of Technology (Caltech) and the University of Arizona have released the largest data set ever

Rensselaer

NATIONAL SCIENCE FOUNDATION

RDA
RESEARCH DATA ALLIANCE

World-wide national and professional communities focusing on 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>

Policy Analysis: Investment, Policy

PUBLISHED: 18 SEPTEMBER 2013

Promote data sharing to advance global research say policy leaders

Richard L. Hudson, Science|Business

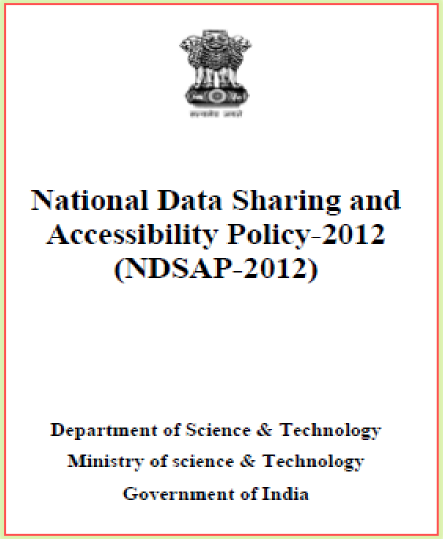
EU and US experts see big benefits from scientists sharing more data - but say global agreement on privacy, literacy and other issues is needed

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

Science, Humanities, Arts Communities

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

RESEARCH DATA ALLIANCE

Fran Berman

The Research Data Alliance (RDA)

- Global community-driven organization launched in March 2013 to accelerate data-driven innovation

- RDA focus is on building the **social, organizational and technical infrastructure** to
  - *reduce barriers to **data sharing and exchange***
  - *accelerate the development of coordinated global data infrastructure*

CREATE → ADOPT → USE

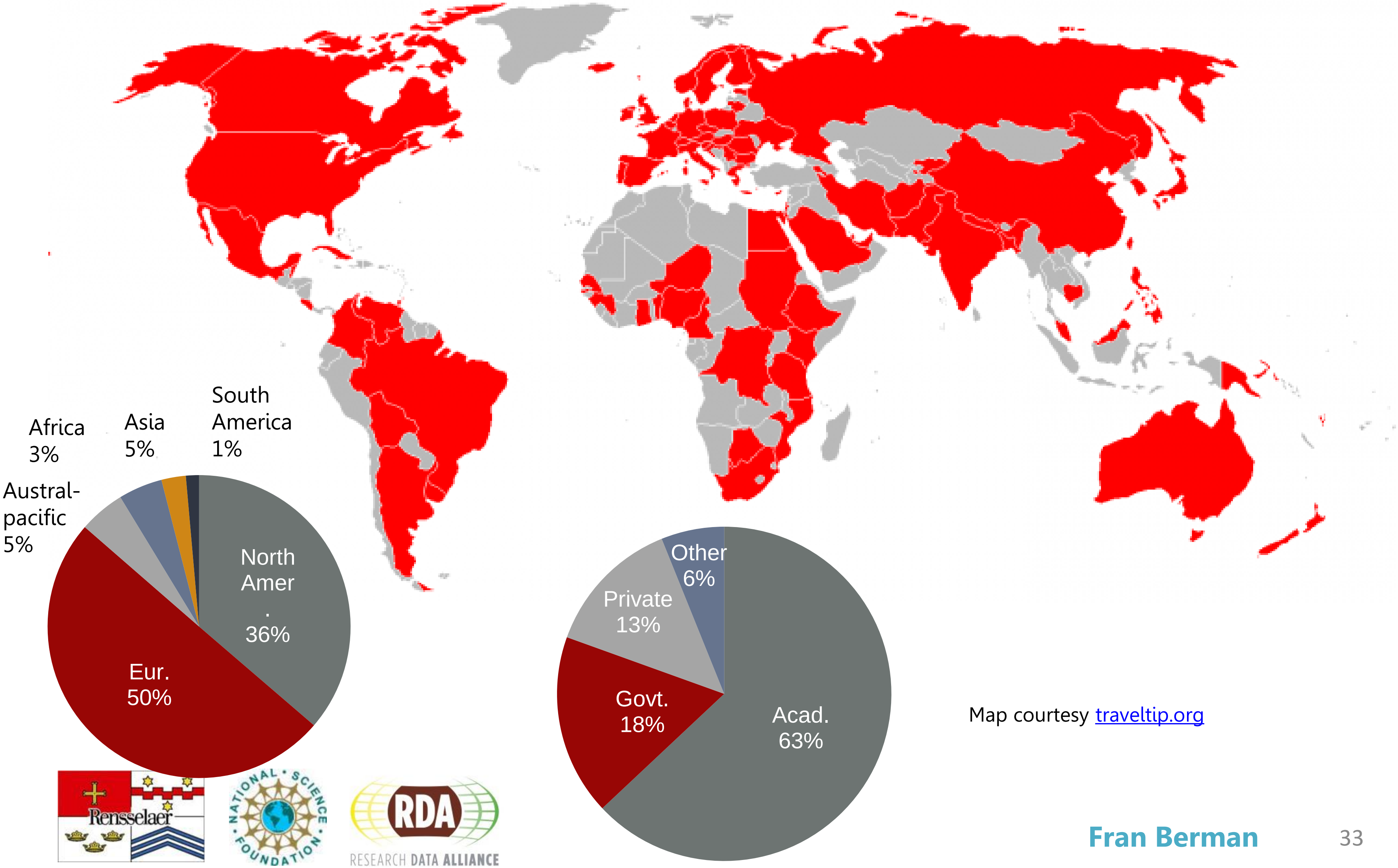
RDA Members come together as

- **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 2300 members from 96 countries (as of 9/14)



Map courtesy traveltip.org

Precipitous growth

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 Organizational Partner Meet-up

First BOFs

380 participants from 22 countries

RDA Second Plenary

September 2013

First Organizational Assembly

6 co-located events

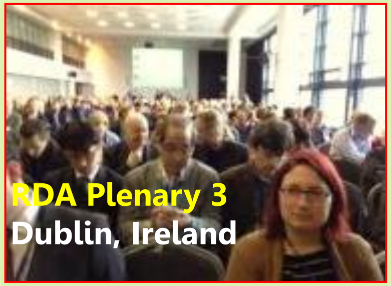
14 BOF, 12 Working Groups, 22 Interest Groups

497 participants

RDA Third Plenary

March 2014

First Working Group exchange meeting



RDA Fourth Plenary

September 2014

RDA Interest (IG) and Working Groups (WG) by focus 1 (as of 9/14)

* under review

Domain Science - focused

- Toxicogenomics Interoperability IG
- Structural Biology IG
- Biodiversity Data Integration IG
- Agricultural Data Interoperability IG
- Wheat Data Interoperability WG
- Digital Practices in History and Ethnography IG
- Geospatial IG
- Marine Data Harmonization IG
- Metabolomics IG
- RDA/CODATA Materials Data Infrastructure and Interoperability IG
- Research Data Needs of the Photon and Neutron Science Community IG
- Defining Urban Data Exchange for Science IG*
- The BioSharing Registry: Connecting data policies, standards and databases in the life sciences WG*
- Urban Quality of Life Indicators WG*

Community Needs - focused

- Community Capability Model IG
- Engagement IG
- RDA / CODATA Summer Schools in Data Science and Cloud Computing in the Developing World WG*
- Development of Cloud Computing Capacity and Education in Developing World Research IG
- Data for Development IG
- Education and Training on handling of research data IG

RDA Interest (IG) and Working Groups (WG) by focus 2 (as of 9/14)

* under review

Reference and Sharing - focused

- Data Citation WG
- Standardization of Data Categories and Codes WG
- RDA/CODATA Legal Interoperability IG
- Reproducibility IG*
- Data Description Registry Interoperability Working Group
- RDA / WDS Publishing Data Bibliometrics WG

Data Stewardship and Services - focused

- Research Data Provenance IG
- Preservation e-infrastructure IG
- RDA / WDS Publishing Data Services WG
- RDA / WDS Publishing Data Workflows WG
- Long-tail of Research Data IG
- RDA/WDS Publishing Data IG
- RDA/WDS Repository Audit and Certification WG
- Domain Repositories Interest Group
- Brokering Interest Group
- ELIXIR Bridging Force IG*
- Libraries for Research Data IG*
- RDA / WDS Certification of Digital Repositories IG
- RDA / WDS Publishing Data Cost Recovery for Data Centres IG

Base Infrastructure - focused

- Data Foundation and Terminology WG
- Metadata Standards Directory WG
- Practical Policy WG
- PID Information Types WG
- Data Type Registries WG
- Data in Context IG
- Big Data Analytics IG
- Data Brokering WG*
- Federated Identity Management IG
- Metadata IG
- PID Interest Group
- Service Management IG
- Data Fabric IG

Organizations committed to joining RDA

- **Organizational Members:**
  - Alliance for Permanent Access
  - American University Library
  - Australian National Data Service
  - Barcelona Supercomputing Center - Centro Nacional de Supercomputación
  - Columbia University Library
  - CNRI
  - CSC
  - Digital Curation Center
  - EIROForum IT Working Group
  - eResearch Services and Scholarly Application Development Division of Information Services, Griffith University
  - European Data Infrastructure (EUDAT)
  - National Institute of Advanced Industrial Science and Technology (AIST), Japan
  - International Association of STM Publishers
  - Internet2
  - Microsoft Research
  - NZ eScience Infrastructure
  - Purdue University Libraries
  - Research Data Canada
  - Scholarly Publishing and Academic Resources Coalition (SPARC)
  - Washington University in St. Louis Libraries
  - Science and Technology Facilities Council

- **Affiliates**
  - CODATA
  - ICSU World Data System
  - ORCID
  - DataCite
  - Global Alliance for Genomics and Health
  - CASRAI

Accelerating data sharing infrastructure, coalescing culture: 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

Partnership with Organizations

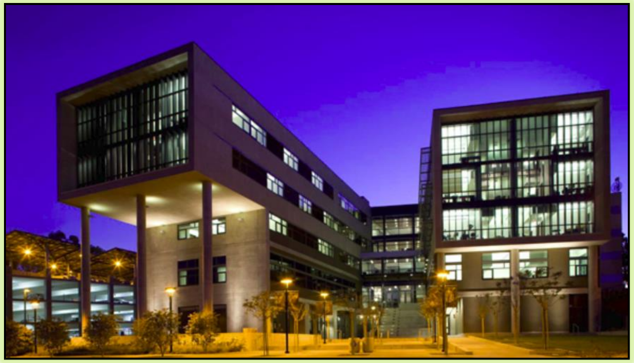
New partners in the Organizational Assembly

Focused strategy to support development of industry infrastructure for data sharing

2015 Focus: Development and adoption of first RDA deliverables

- **March 9-11, 2015: RDA/US hosting RDA Plenary 5 in San Diego**
  - Working Meeting of the RDA with co-located community meetings

- **San Diego Supercomputer Center hosting RDA Adoption Day on March 8**



Thank You

ECONOMICS

Budget realistically for the costs of data stewardship and preservation

Prioritize the "data bill" at the same level as other critical infrastructure.

ECOLOGY

Create and implement a **data management and stewardship plan** for your project for a reasonable fixed term of time.

Make your data available to the community (as appropriate) by curating it and ingesting it into a publicly accessible repository

POLITICS

Adopt / support policy and practice that enables the development and continued maintenance of sustainable stewardship, data sharing, and broad access

CULTURE

Contribute /create a local / community culture of **data sharing**

Cite and publish your data when you write about your results. Work with your professional societies and conferences to include **"data sessions"** *(idea from Sibel Adali)*