



Data Management Plans Through the Eyes of the United States National Science Foundation

Patricia Knezek, Ph.D.

Senior Advisor

Division of Advanced Cyberinfrastructure

Directorate for Computer & Information Science & Engineering

National Science Foundation

Summary and Objectives

- The National Science Foundation (NSF): Who we are
- Overview of the current Public Access and Data Management Plan requirements
- Challenges and Considerations for (A)DMPs
- Next steps

NSF Mission



Characteristics of NSF: Ubiquity, Urgency, and Engagement

Ubiquity

Advances in science and engineering are permeating the way we work, communicate, learn, and discover.

Urgency

NSF research and education are rapidly evolving and accelerating the pace of discovery and innovation, with profound societal and economic impact.

Engagement

The key strength and asset of NSF is the scientific community and the general public and their engagement.





NATIONAL SCIENCE FOUNDATION

NATIONAL SCIENCE BOARD (NSB)

Dan E. Arvizu
Chair

Kelvin K. Droegemeier
Vice Chair
703.292.7000

NATIONAL SCIENCE BOARD OFFICE

Michael Van Woert
Executive Officer
703.292.7000

OFFICE OF INSPECTOR GENERAL (OIG)

Allison C. Lerner, Inspector General
703.292.7100

Richard Buckius
Chief Operating Officer

OFFICE OF THE DIRECTOR
703.292.8000

France A. Córdoba
Director

Vacant
Deputy Director

OFFICE OF DIVERSITY & INCLUSION (ODI)

Rhonda Davis, Head
703.292.8020

OFFICE OF THE GENERAL COUNSEL (OGC)

Lawrence Rudolph, General Counsel
Peggy Hoyte, Deputy GC
703.292.8060

OFFICE OF INTEGRATIVE ACTIVITIES (OIA)

Suzanne Iacono, Head
703.292.8040

OFFICE OF INTERNATIONAL SCIENCE & ENGINEERING (OISE)

Rebecca Keiser, Head
703.292.8710

OFFICE OF LEGISLATIVE & PUBLIC AFFAIRS (OLPA)

Amanda Greenwell, Head
703.292.8070

DIRECTORATE FOR BIOLOGICAL SCIENCES (BIO)

James L. Oids, Assistant Director
Jane Silverthorne, Deputy AD
703.292.8400

DIRECTORATE FOR COMPUTER & INFORMATION SCIENCE & ENGINEERING (CISE)

James F. Kurose, Assistant Director
Erwin Gianchandani, Acting Deputy AD
703.292.8900

DIRECTORATE FOR EDUCATION & HUMAN RESOURCES (EHR)

Joan Ferrini-Mundy, Assistant Director
William (Jim) Lewis, Deputy AD
703.292.8600

DIRECTORATE FOR ENGINEERING (ENG)

Pramod P. Khargonekar, Assistant Director
Grace Wang, Deputy AD
703.292.8300

DIRECTORATE FOR GEOSCIENCES (GEO)

Roger Wakimoto, Assistant Director
Margaret Cavanaugh, Deputy AD
703.292.8500

DIRECTORATE FOR MATHEMATICAL & PHYSICAL SCIENCES (MPS)

Fleming Crim, Assistant Director
Clifford Gabriel, Acting Deputy AD
703.292.8800

DIRECTORATE FOR SOCIAL, BEHAVIORAL, & ECONOMIC SCIENCES (SBE)

Fay L. Cook, Assistant Director
Kellina M. Craig-Henderson, Deputy AD
703.292.8700

OFFICE OF BUDGET, FINANCE, & AWARD MANAGEMENT (BFA)

Martha A. Rubenstein, Head / Chief Financial Officer
Teresa Granovirtz, Deputy Head
703.292.8200

OFFICE OF INFORMATION & RESOURCE MANAGEMENT (IRM)

Jeanne Head / Capital Officer
Donna Depuy
703.292.8100

DIVISION OF BIOLOGICAL INFRASTRUCTURE (DBI)

Muriel E. Roston, Division Director
703.292.8470

DIVISION OF COMPUTER & NETWORK SYSTEMS (CNS)

Peter Arzberger, Acting Division Director
703.292.8950

DIVISION OF GRADUATE EDUCATION (DGE)

Dean Cavallaro, Division Director
703.292.8630

DIVISION OF CHEMICAL, BIOENGINEERING, ENVIRONMENTAL & TRANSPORT SYSTEMS (CBET)

JoAnn Lighty, Division Director
703.292.8320

DIVISION OF ATMOSPHERIC & GEOSPACE SCIENCES (AGS)

Paul Shapiro, Division Director
703.292.8520

DIVISION OF ASTRONOMICAL SCIENCES (AST)

James Ulvestad, Division Director
703.292.8620

DIVISION OF BEHAVIORAL & COGNITIVE SCIENCES (BCS)

Howard Nusbaum, Division Director
703.292.8740

BUDGET DIVISION (BD)

Michael Sieverts, Division Director
703.292.8260

DIVISION OF ADMINISTRATIVE SERVICES (DAS)

Wonzie Gardner, Acting Division Director
703.292.8180

DIVISION OF ENVIRONMENTAL BIOLOGY (DEB)

Paula M. Mabey, Division Director
703.292.8490

DIVISION OF COMPUTING & COMMUNICATION FOUNDATIONS (CCF)

Rao Kosaraju, Division Director
703.292.8910

DIVISION OF HUMAN RESOURCE DEVELOPMENT (HRD)

Sylvia James, Division Director
703.292.8640

DIVISION OF CIVIL, MECHANICAL & MANUFACTURING INNOVATION (CMMI)

Deborah Goodings, Division Director
703.292.8360

DIVISION OF EARTH SCIENCES (EAR)

Carol Frost, Division Director
703.292.8590

DIVISION OF CHEMISTRY (CHE)

Carol Beisel, Acting Division Director
703.292.8840

DIVISION OF SOCIAL & ECONOMIC SCIENCES (SES)

Alan Tomkins, Acting Division Director
703.292.8790

DIVISION OF ACQUISITION AND COOPERATIVE SUPPORT (DACS)

Jeffery Lupis, Division Director
703.292.8240

DIVISION OF INFORMATION SYSTEMS (DIS)

Dorothy Aronson, Division Director
703.292.8150

DIVISION OF INTEGRATIVE ORGANISMAL SYSTEMS (IOS)

Heinz O. de Couet, Division Director
703.292.8420

DIVISION OF ADVANCED CYBERINFRASTRUCTURE (ACI)

Irene Quatters, Division Director
703.292.8970

DIVISION OF RESEARCH ON LEARNING IN FORMAL & INFORMAL SETTINGS (DRL)

Evan Holt, Division Director
703.292.8620

DIVISION OF ELECTRICAL, COMMUNICATIONS & CYBER SYSTEMS (ECCS)

Samir El-Ghazally, Division Director
703.292.8330

DIVISION OF OCEAN SCIENCES (OCS)

Richard Murray, Division Director
703.292.8500

DIVISION OF MATERIALS RESEARCH (DMR)

Linda S. Sapochak, Acting Division Director
703.292.8810

NATIONAL CENTER FOR SCIENCE AND ENGINEERING STATISTICS (NCSES)

John Gawell, Division Director
703.292.8790

DIVISION OF FINANCIAL MANAGEMENT (DFM)

Michael Wetlow, Division Director
703.292.8290

DIVISION OF HUMAN RESOURCE MANAGEMENT (HRM)

Judy Sunley, Division Director
703.292.8190

DIVISION OF MOLECULAR & CELLULAR BIOSCIENCES (MCB)

Linda E. Hyman, Division Director
703.292.8440

DIVISION OF INFORMATION & INTELLIGENT SYSTEMS (IIS)

Lynne E. Parker, Division Director
703.292.8930

DIVISION OF UNDERGRADUATE EDUCATION (DUE)

Susan Singer, Division Director
703.292.8670

DIVISION OF ENGINEERING EDUCATION & CENTERS (EEC)

Mario Robles, Division Director
703.292.8380

DIVISION OF POLAR PROGRAMS (PLR)

Kelly Falkner, Division Director
703.292.8030

DIVISION OF MATHEMATICAL SCIENCES (DMS)

Michael Vogelius, Division Director
703.292.8670

DIVISION OF PHYSICS (PHY)

Denise Caldwell, Division Director
703.292.8890

DIVISION OF INSTITUTION & AWARD SUPPORT (DIAS)

Dale Bell, Division Director
703.292.8230

DIVISION OF HUMAN RESOURCE MANAGEMENT (HRM)

Judy Sunley, Division Director
703.292.8190

OFFICE OF EMERGING FRONTIERS (EF)

Charles Lianke, Program Director
703.292.8508

DIVISION OF INDUSTRIAL INNOVATION & PARTNERSHIPS (IP)

Barry Johnson, Division Director
703.292.8090

OFFICE OF EMERGING FRONTIERS IN RESEARCH & INNOVATION (EFRRI)

Sohi Rashleger, Senior Advisor
703.292.8301

DIVISION OF INDUSTRIAL INNOVATION & PARTNERSHIPS (IP)

Barry Johnson, Division Director
703.292.8090

OFFICE OF MULTIDISCIPLINARY ACTIVITIES (OMA)

Clark Cooper, Office Head
703.292.8680

OFFICE OF MULTIDISCIPLINARY ACTIVITIES (OMA)

Clark Cooper, Office Head
703.292.8680

LARGE FACILITIES OFFICE

Matthew J. Hawkins, Deputy Director
703.292.4416

LARGE FACILITIES OFFICE

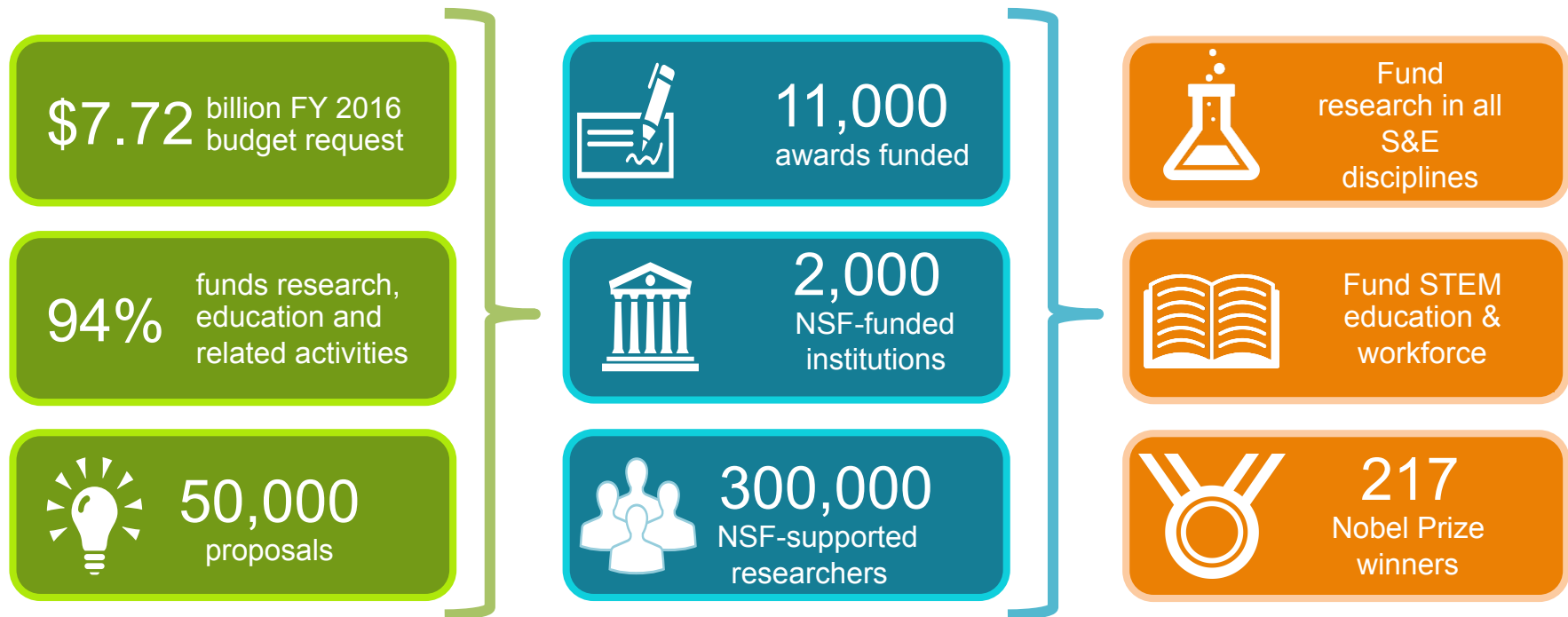
Matthew J. Hawkins, Deputy Director
703.292.4416

LARGE FACILITIES OFFICE

Matthew J. Hawkins, Deputy Director
703.292.4416

National Science Foundation
4201 Wilson Boulevard
Arlington, Virginia 22230
TEL: 703.292.5111 | FIRS: 800.877.8339 | TDD: 800.281.8749

NSF by the Numbers



U. S. Federal Policy on Public Access

- U.S. Office of Science and Technology Policy (OSTP) issued a memo in March 2013.
 - directed federal agencies to develop plans to make *publicly available* to the “*greatest extent and with the fewest constraints possible and consistent with law*” the “*direct results of federally funded scientific research.*”

NSF Public Access

- Clear and open communication of research results is central to the progress of science
 - Publications
 - Data (full range of research products)



[NSF] “continues its commitment to expand public access to the results of its funded research. Public access is intended to accelerate the dissemination of fundamental research results that will advance the frontiers of knowledge and help ensure the nation’s future prosperity”

F. Cordova, Director, NSF, 3/18/15

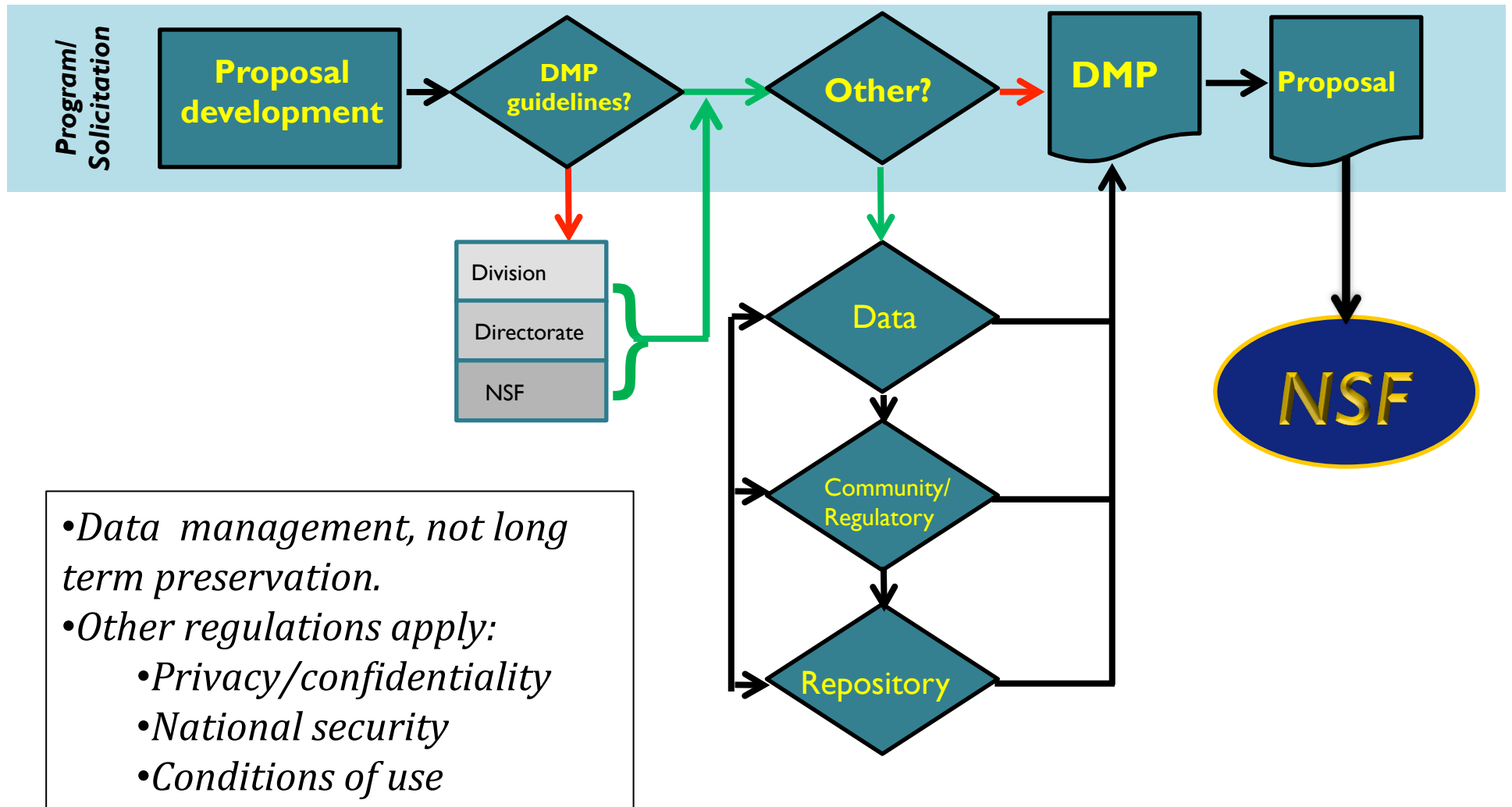
NSF Response to OSTP Memo

- NSF developed a plan that is consistent with OSTP policy objectives
 - Publication deposit in support of public access is a **new requirement**
 - Deposit is separate, but integrated with reporting in Research.gov to minimize burden
 - Builds on **existing** reporting requirements
 - Article Processing charges can be requested as a direct cost, as is **current policy**
 - Current DMP policies **remain in place** with activities intended to enable consistent identification, description, and management through the directorates/communities

NFS Public Access Plan: data management approach

Key points:

- “data management is dynamic and practices vary substantially across the broad range of scientific disciplines supported by NSF” [NSF 15-52]
- “What constitutes reasonable data management and access will be determined by the community of interest *through the process of peer review and program management.*” [Data Management & Sharing Frequently Asked Questions (FAQs)]
- bottom-up implementation in context of top-down guiding principles
 - “one size” does not fit all of science, engineering and education



Guidelines are community driven

NSF's Data Sharing Policy covers: (I)

- **Publication**

a. Investigators are expected to promptly prepare and submit for publication, with authorship that accurately reflects the contributions of those involved, all significant findings from work conducted under NSF grants. Grantees are expected to permit and encourage such publication by those actually performing that work, unless a grantee intends to publish or disseminate such findings itself.

- **Data (independent of format)**

b. Investigators are expected to share with other researchers, at no more than incremental cost, within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. Privileged or confidential information should be released in a form that protects the privacy of individuals and subjects involved. General adjustments and, where essential, exceptions to this sharing expectation may be specified by the funding NSF Program Division/Office for a particular field or discipline to safeguard the rights of individuals and subjects, the validity of results, or the integrity of collections or to accommodate the legitimate interests of investigators. A grantee or investigator also may request a particular adjustment or exception from the cognizant NSF Program Officer.

NSF's Data Sharing Policy covers: (2)

- ***Software and inventions***

c. Investigators and grantees are encouraged to share software and inventions created under the grant or otherwise make them or their products widely available and usable.

- ***Intellectual property***

d. NSF normally allows grantees to retain principal legal rights to intellectual property developed under NSF grants to provide incentives for development and dissemination of inventions, software, and publications that can enhance their usefulness, accessibility and upkeep. Such incentives do not, however, reduce the responsibility that investigators and organizations have as members of the scientific and engineering community, to make results, data and collections available to other researchers.

- ***Decentralized administration***

e. NSF program management will implement these policies for dissemination and sharing of research results, in a way appropriate to field and circumstances, through the proposal review process; through award negotiations and conditions; and through appropriate support and incentives for data collection, documentation, dissemination, storage and the like.

NSF's current data management plan:

- Is a required element of all proposals (two-page limit)
- Describes how the PI will conform to data policy associated with the competition to which the proposal submitted. If none exists, to NSF's generic policy
- Is evaluated during merit review as Intellectual Merit, Broader Impacts or both, as appropriate
- May simply say "No plan is needed", but this needs a justification. Statement will be subject to peer and Program Officer review/consideration

DMP may include:

- **Types of data**, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
- **Standards** to be used for data and metadata format and content (document where existing standards are absent or deemed inadequate, along with any proposed solutions or remedies);
- **Policies for access and sharing** including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- **Policies and provisions for re-use**, re-distribution, and the production of derivatives; and
- **Plans for archiving** data, samples, and other research products, and for preservation of access to them.

Challenges for DMP development

- Trade-offs between ease of use and protections
- Provenance
 - Versioning of software, workflows, and data
 - IP difficult to establish
 - Future uses difficult to define
- New data, new risks
 - Social media as a source of data
 - “Trace” data

More considerations for DMPs

- Specific attention to the physical disposition of the data
 - Where are the data to be deposited?
 - Are the basic security controls in place
 - What restrictions, if any, may be imposed? For how long? And for whom?
- Willingness to differentiate between the “raw” data (which may be restricted, as appropriate) and the models and techniques (which can be released)

DMP: Where do we go from here?

- Individual directorates undertaking self-studies in anticipation of providing updated DMP guidance
 - What do investigators actually do?
 - Is the DMP effective?
- Under discussion:
 - foundation-wide workshop
 - pilots using persistent identifiers
- Longer term:
 - exploit DMPs for purpose of compliance (A)DMP
 - consultation with multiple agencies, evolve DMPs as experience is gained, technology changes
 - understand roles, responsibilities, business models

Summary and Future Perspective

- NSF continues its deep commitment to expand public access to results of its funded research
 - NSF Public Access Plan: publications, data products
- Realizing the enormous potential of Data requires a long term, bold, sustainable, and comprehensive approach, by and by our partners across the globe
 - Advancing science is truly a borderless enterprise
 - Meetings like this workshop needed to build bridges (understanding, people) to advance science, reap societal benefits

STATISTICAL,
COMPUTATIONAL
FOUNDATIONS
INFERENC
SEMANTICS
EHR
ANALYTICS
PRIVA
PUBLIC
ACCESS
ENG
DISCOVI
EDUCATION
WORKFORCE
DATA MINING

DATA SCIENCE

FUNDAMENTAL RESEARCH
CISE
GEO
MACHINE
LEARNING
CYBERSECURITY
SBE
BIO
DOMAIN
SCIENCE
CHALLENGES
SYSTEMS ARCHITECTURE
INTERNET
PRODUCABILITY
STATISTICS
RESEARCH
DATA
CYBERINFRASTRUCTURE
HUMAN-DATA INTERFACE
MPS
CAUSALITY
VISUALIZATION

Thank you

- Back up slides

A CI Architectural Reference Model

↑
Increasing disciplinary emphasis



Science APIs, portals,
gateways



Science APIs, portals,
gateways

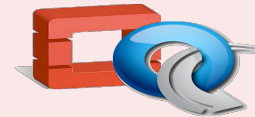
Science APIs, portals,
gateways

iRODS

federated storage



HPC access, community



open system software

single
sign-on

international

private, commercial cloud campus, national resources

**NSF-supported CI
ecosystem**



National/International Research and Education Network

NSF “Big Ideas”

Science AAAS

Authors | Members | Librarians | Advertisers

Home News Journals Topics Careers Search

Latest News ScienceInsider ScienceShots Sifter From the Magazine About News Quizzes



Better understanding the changing Arctic is one item on a new list of big ideas that should shape the National Science Foundation's work.

NASA/Kathryn Hansen

NSF director unveils big ideas, with an eye on the next president and Congress

By **Jeffrey Mervis** | May. 10, 2016 , 3:30 PM

NSF “Big Ideas”

RESEARCH IDEAS

- Harnessing Data for 21st Century Science and Engineering
- Shaping the new Human – Technology Frontier
- Understanding the Rules of Life: Predicting Phenotype
- The Quantum Leap: Leading the Next Quantum Revolution
- Navigating the New Arctic
- Windows on the Universe: The Era of Multi-messenger Astrophysics

PROCESS IDEAS

- Growing Convergent Research at NSF
- Mid-scale Research Infrastructure
- NSF 2050

*Video of NSF presentation and discussion is at:

http://www.tvworldwide.com/events/nsf/160505/globe_show/default_go_archive.cfm?gsid=2957&type=flv&test=0&live

(the presentation/discussion starts about 20 minutes into this video)

Harnessing the Data Revolution

Embodiment of innovations in robust, comprehensive, open, science-driven, CI ecosystems

accelerating data-intensive research, including large-scale facilities

fundamental research: mathematics, statistics, computer & computational science

fundamental research: algorithms, systems

data discovery, integration; predictive analytics, data mining, machine learning; data semantics; open data-centric architectures, systems; data integrity, access; benchmark data sets; privacy, human-data interface

Data-intensive domain research:

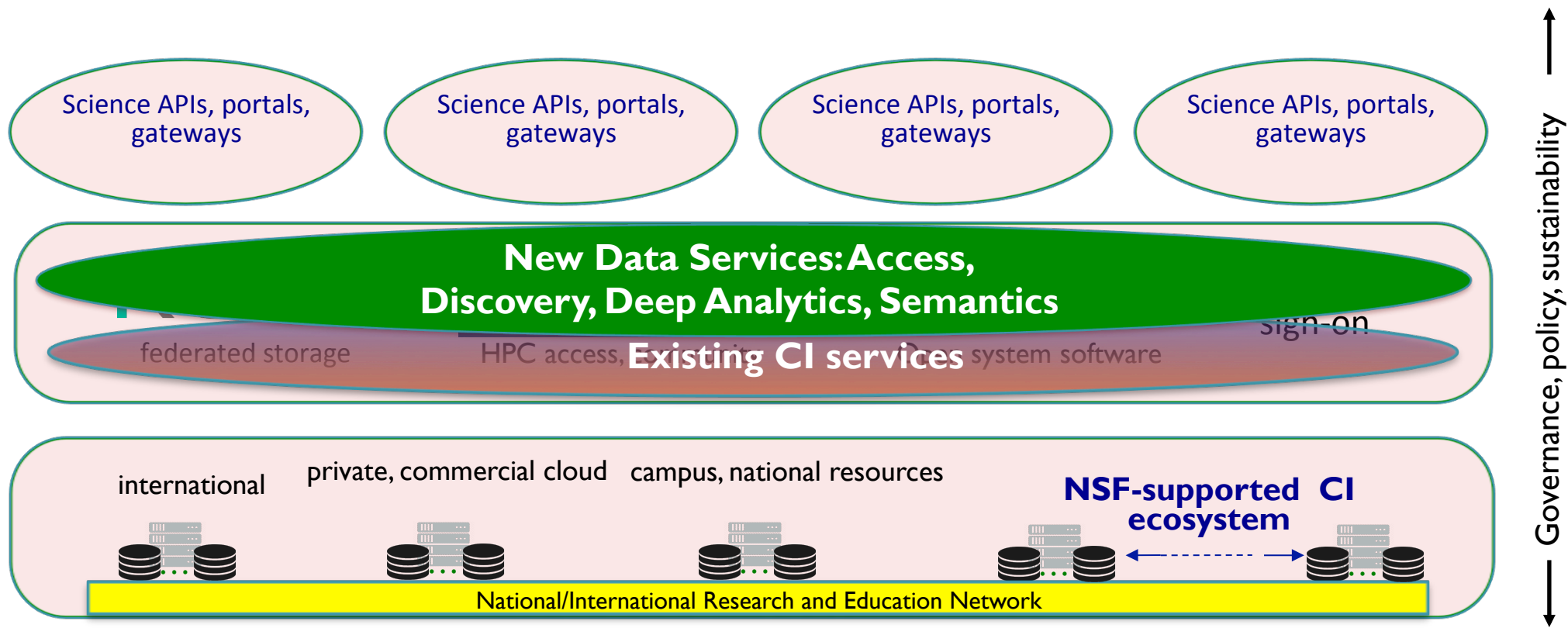
use advances in data science and CI to further research

Development, evaluation of innovative learning opportunities, educational pathways: grounded in an

education-research-based framework

A vision for research cyberinfrastructure

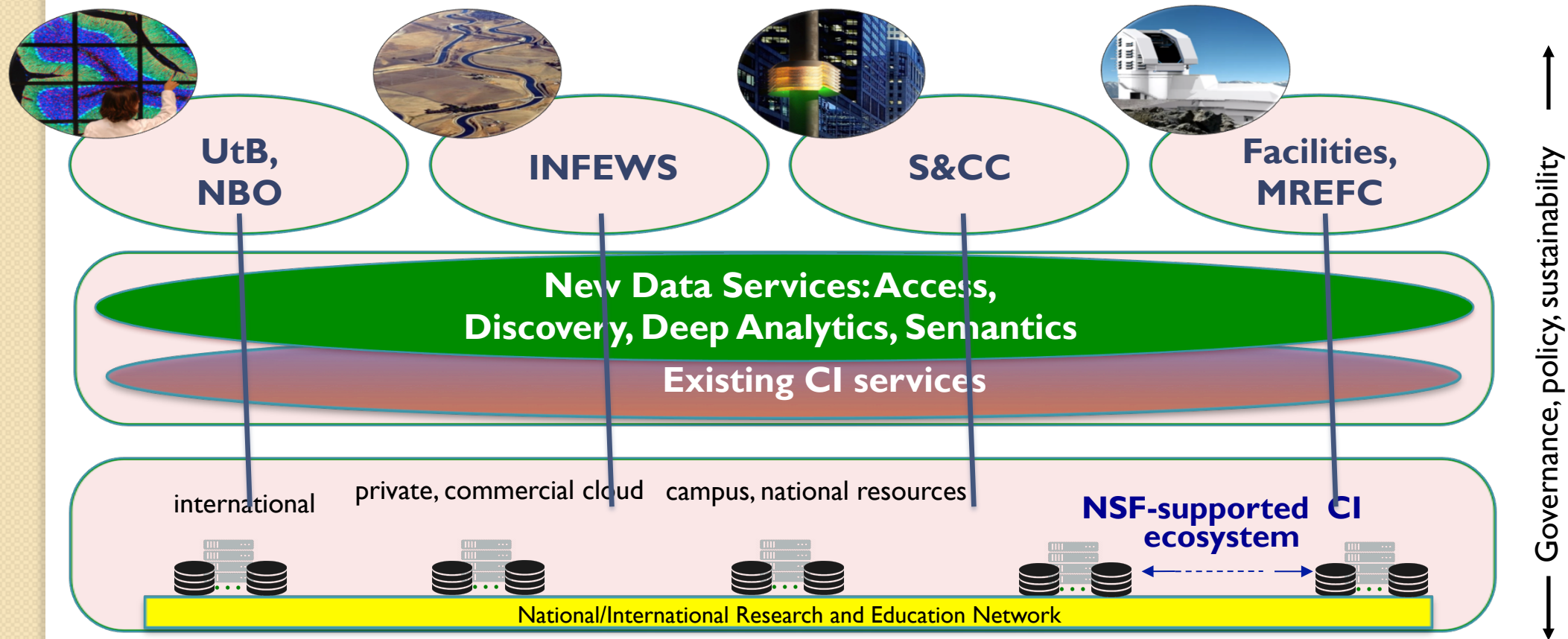
Architecting an open national data infrastructure



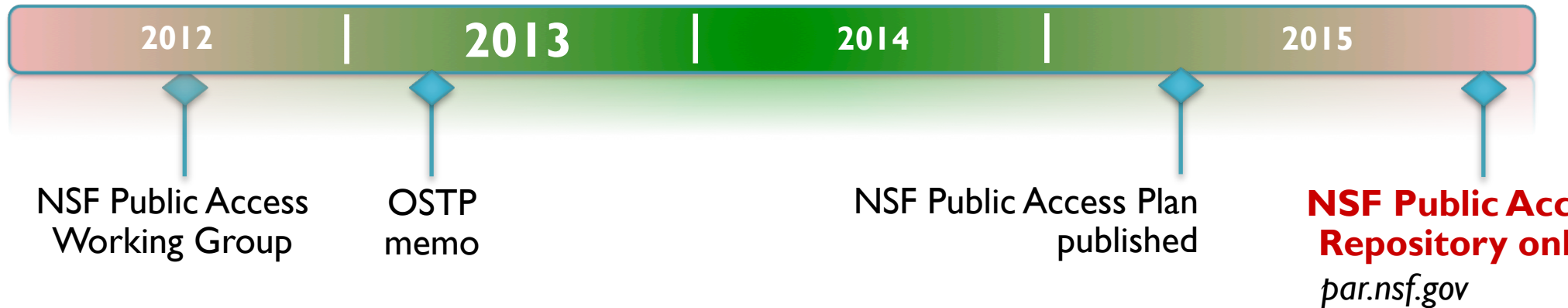
A vision for research cyberinfrastructure

Architecting an open national data infrastructure

Enabling and accelerating science drivers, including NSF initiatives & facilities

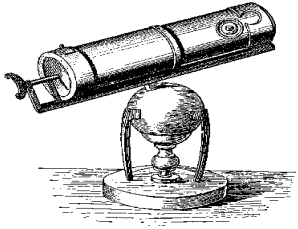


NSF Public Access Repository



- Provide public access to journal, juried conference papers
- Leverage existing systems, workflows:
 - Integrated with NSF-internal proposal, award management
 - Leverage DOE/OSTI infrastructure for publications; publisher/library services (e.g. CrossRef)
 - Extensible: other research products, federation
- PI's must deposit publications in PAR, awards made FY 2016 onward

Evolving research paradigms



$$\oint \mathbf{E} \cdot d\mathbf{A} = \frac{q_{enc}}{\epsilon_0}$$

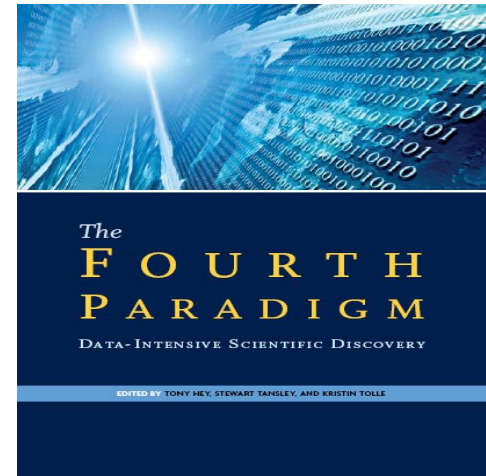
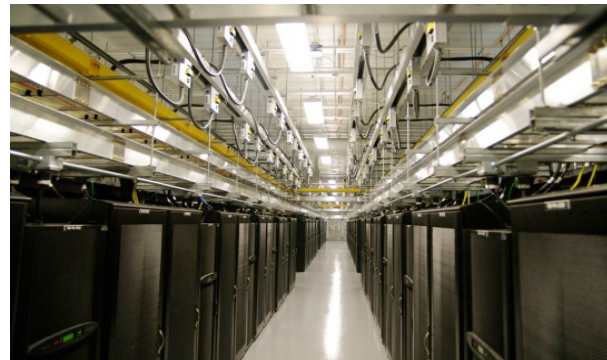
$$\oint \mathbf{B} \cdot d\mathbf{A} = 0$$

$$\oint \mathbf{E} \cdot d\mathbf{s} = -\frac{d\Phi_B}{dt}$$

$$\oint \mathbf{B} \cdot d\mathbf{s} = \mu_0 \epsilon_0 \frac{d\Phi_E}{dt} + \mu_0 i_{enc}$$

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u_i)}{\partial x_i} = S_m$$

$$\frac{\partial (\rho u_i)}{\partial t} + \frac{\partial (\rho u_i u_j)}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j} + \rho g_i + F_i$$



Experimental

Theoretical

Computational

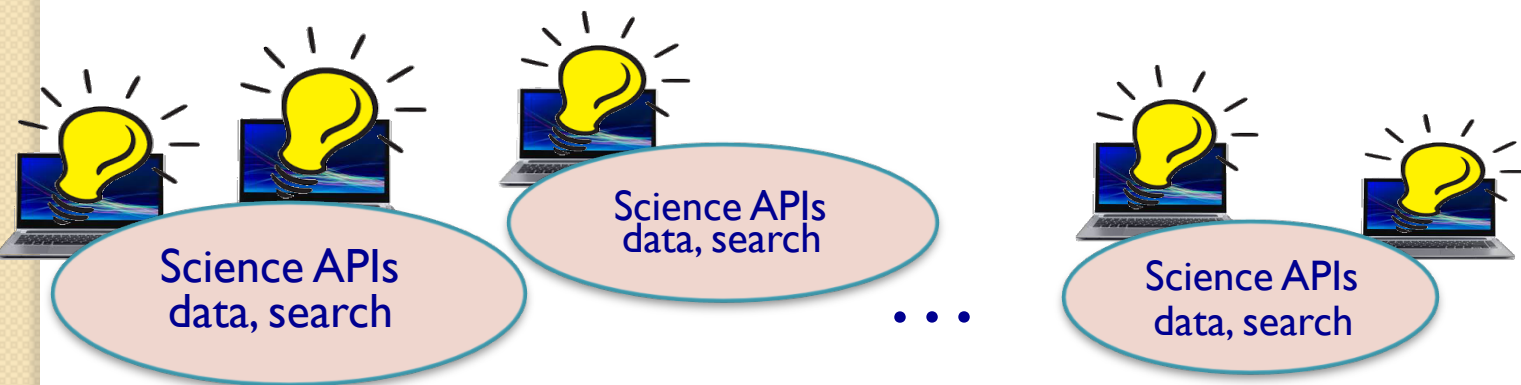
Data

The Emerging Data Landscape

- Data and the domains:
 - Domains: science, engineering, social science, education, business, finance, ...
 - Systematic approaches to data management, curation
 - Domain-specific data-analytic techniques
 - Need for cyberinfrastructure
- Data Science as a discipline:
 - Application domains, computing, statistics, policy

NSF supports Research Cyberinfrastructure to Enable Scientific Collaboration and Discovery

Data-sharing requires significant investments at many levels

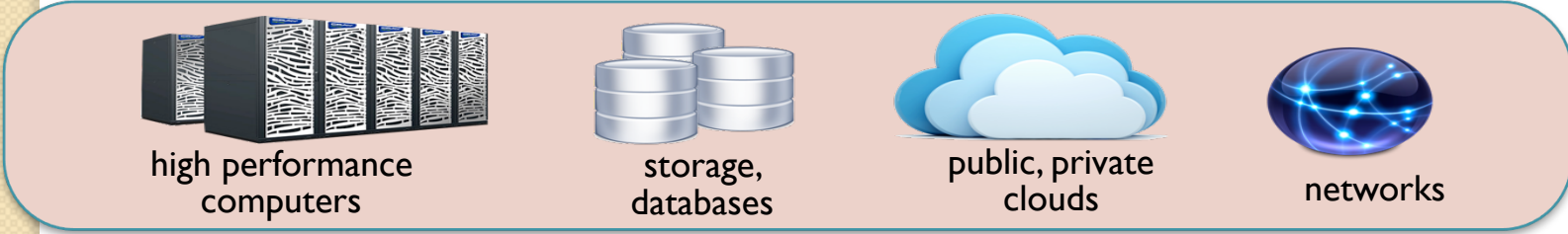


science and discovery environment

domain-specific services



low-level software services



hardware resources

Federal Agency Activities, Programs

- **NSF:** BIGDATA , Computational and Data-Enabled Science and Engineering (CDS&E), Data Infrastructure Building Blocks (DIBB), NSF Research Traineeships (NRT)
- **NIH:** Big Data to Knowledge (BD2K)
- **DARPA:** Information Innovation Office (I2O)
- **NIST:** Big Data Public Working Group
- **DOE:** Scientific Discovery through Advanced Computing (SciDA)

- **Networking and Information Technology Research and Development (NITRD) Program:** Big Data Senior Steering Group
 - Membership: 18 federal R&D agencies
 - Co-chaired by NSF, NIH

NSF Data-related activities

FOUNDATIONAL RESEARCH:

CDS&E
BIGDATA
III core program

CYBERINFRASTRUCTURE:

DIBBS
Wrangler, Comet ,Jetstream
CC*DNI

EDUCATION & WORKFORCE DEVELOPMENT

NRT

COMMUNITY BUILDING

White House BD Partners WS
Data to Knowledge to Action
BD Strategic Initiatives WS
BD Regional Innovation Hubs
EarthCube
RDA

Data for Scientific Discovery & Action (D4SDA)

*Enabling 21st century science, engineering, and education to move toward effective use of digital data
advance discovery*

Announced in FY 2017 budget request:

- Promote foundational research in critical techniques, technologies
- Provide innovative, reusable data and knowledge infrastructure to support data-intensive science
- Enable/incent science community to address data governance, lifecycle issues
- Educate data-savvy workforce of scientists, engineers, educators

RDA/US



- building the social, technical bridges that enable open sharing of data
- “grass roots”, “bottom-up” community effort towards global data sharing
 - RDA-USA is supported by NSF under Grant No. I349002, and the National Institute of Standards and Technology (NIST)
- RDA/US is part of the larger, global community-driven RDA
 - international collaboration, participation, perspectives are critical

What does the plan say?

(consistent with OSTP policy objectives)

- Requires deposit of journal articles and juried conference papers funded by awards resulting from proposals submitted in January 2016 (PAPPG) to be made publicly available no later than 12 months after publication in the NSF Public Access Repository, hosted by DOE/OSTI
- Allows for a waiver to the 12-month embargo
- Retains current DMP requirement, allowance for costs, and data citation and calls for community engagement to support consistent data management best practice
- Supports public search through:
 - Existing award search mechanisms
 - Search capability on the NSF Public Access Repository, hosted by DOE/OSTI
 - Expose metadata to third party search systems (future)
- Leverages current programs, policies, and systems
- Calls for a Working Group to provide oversight
- Establishes a website with the Plan, opportunity for feedback, and FAQs
- Calls for regular updates to the NSB and OSTP/OMB
- Allows for evolution to other products of NSF-funded research