



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

# Research Data Management

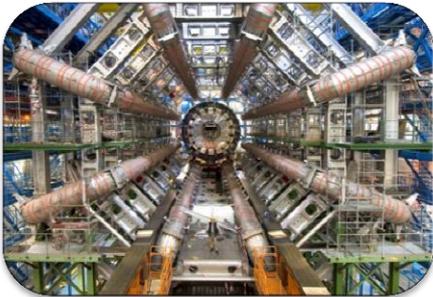
NITRD FASTER

July 18, 2016

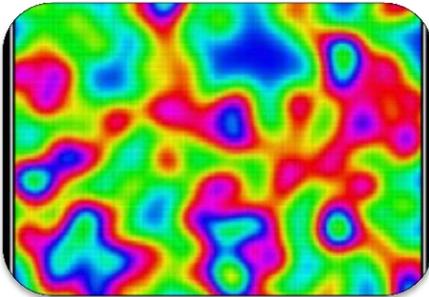
Laura Biven, PhD  
Senior Science and Technology Advisor  
Office of the Deputy Director for Science Programs (SC-2)  
U.S. Department of Energy  
[Laura.Biven@science.doe.gov](mailto:Laura.Biven@science.doe.gov)

# Office of Science FY 2016: \$5.35B

Dr. Murray  
slide



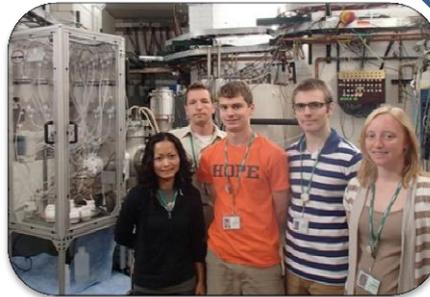
Largest Supporter of Physical Sciences in the U.S.\*



Research: 42%, \$2.2B



~40% of Research to Universities



> 22,000 Scientists Supported



Funding at >300 Institutions including all 17 DOE Labs



Construction: 13.5%, \$723M



Facility Operations: 38%, \$2.02B



>33,000 Scientific Facility Users\*\*

\* 43% of all physical sciences, 30% of computer science and math

\*\* from all 50 states and DC



# The Office of Science research portfolio

## Advanced Scientific Computing Research

- Delivering world leading computational and networking capabilities to extend the frontiers of science and technology

## Basic Energy Sciences

- Understanding, predicting, and ultimately controlling matter and energy flow at the electronic, atomic, and molecular levels

## Biological and Environmental Research

- Understanding complex biological, climatic, and environmental systems

## Fusion Energy Sciences

- Building the scientific foundations for a fusion energy source

## High Energy Physics

- Understanding how the universe works at its most fundamental level

## Nuclear Physics

- Discovering, exploring, and understanding all forms of nuclear matter



# The DOE/SC Labs Today



**Berkeley, California**  
 202 acres and 97 buildings  
 3,396 FTEs  
 950 students & postdocs  
 9,320 facility users  
[www.lbl.gov](http://www.lbl.gov)



**Pacific Northwest**  
 NATIONAL LABORATORY



**Richland, Washington**  
 346 acres and 19 buildings  
 4,344 FTEs  
 550 students & postdocs  
 1,733 facility users  
[www.pnnl.gov](http://www.pnnl.gov)



THE Ames Laboratory



**Ames, Iowa**  
 8 acres and 12 buildings  
 308 FTEs  
 158 students & postdocs  
[www.ameslab.gov](http://www.ameslab.gov)



**Batavia, Illinois**  
 6,800 acres and 354 buildings  
 1,720 FTEs  
 55 students & postdocs  
 2,097 facility users  
[www.fnal.gov](http://www.fnal.gov)



**Argonne**  
 NATIONAL LABORATORY



**Argonne, Illinois**  
 1,517 acres and 100 buildings  
 3,460 FTEs  
 1,054 students & postdocs  
 6,547 facility users  
[www.anl.gov](http://www.anl.gov)



**Menlo Park, California**  
 426 acres and 151 buildings  
 1,596 FTEs  
 213 students & postdocs  
 4,474 facility users  
[www.slac.stanford.edu](http://www.slac.stanford.edu)



**Oak Ridge, Tennessee**  
 4,421 acres and 194 buildings  
 4,586 FTEs  
 1,080 students & postdocs  
 3,215 facility users  
[www.ornl.gov](http://www.ornl.gov)



**Newport News, Virginia**  
 169 acres and 72 buildings  
 729 FTEs  
 60 students & postdocs  
 1,261 facility users  
[www.jlab.org](http://www.jlab.org)



PRINCETON  
 PLASMA PHYSICS  
 LABORATORY



**Princeton, New Jersey**  
 89 acres and 34 buildings  
 429 FTEs  
 54 students & postdocs  
 290 facility users  
[www.pppl.gov](http://www.pppl.gov)



**Upton, New York**  
 5,322 acres and 310 buildings  
 2,882 FTEs  
 642 students & postdocs  
 4,134 facility users  
[www.bnl.gov](http://www.bnl.gov)



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
 Science

# Office of Science User Facilities



**28 world-leading facilities serving over 33,000 researchers annually**

- supercomputers,
  - high intensity x-ray, neutron, and electron sources,
  - nanoscience facilities,
  - genomic sequencing facilities,
  - particle accelerators,
  - fusion/plasma physics facilities, and
  - atmospheric monitoring capabilities.
- 
- **Open access; allocation determined through peer review of proposals**
  - **Free for non-proprietary work published in the open literature**
  - Full cost recovery for proprietary work



# FY 2015 User Crossover Analysis

	C-Mod	ALCF	ALS	APS	ARM	ATF	ATLAS	CEBAF	CFN	CINT	CNM	CNMS	DIII-D	EMSL	Esnet	FACET	Fermilab AC	HFIR	JGI	LCLS	NERSC	NSLS-II	NSTX-U	OLCF	RHIC	SNS	SSRL	TMF	
Alcator C-Mod	224	2	1										48								40		35	2					
ALCF	2	990	4	18	9		1	5	2	3	15		3	12	1	1	10	3		4	281		2	169	2	4	3	2	
ALS	1	4	2560	286	13			6	6	9	8	4		18			2	13	2	65	90	5			1	15	223	96	
APS		18	286	5471			8	14	19	8	154	29		19		2	5	77	3	77	26	25	1	1		163	183	5	
ARM		9	13		1121			2						23			2				71			15			3		
ATF						75		1								10					4				2				
ATLAS		1		8			392	5		1	5										3				1				
CEBAF		5	6	14	2	1	5	1510	1	1	7	3		3			36	2		1	44			14	29	7	4	1	
CFN		2	6	19				1	493	1	2	3		1				3		1	19	12		2		3	3	2	
CINT		3	9	8			1	1	1	502		9						2			6	1		4		2	8	2	
CNM		15	8	154			5	7	2		529			2			4	2		4	16	2			1		8	1	
CNMS			4	29				3	3	9		575					1	1	29	1	16			3		74	1	2	
DIII-D	48	3											557				1				86		87	14				1	
EMSL		12	18	19	23			3	1		2			713			2	2	39		64			13		3	12	3	
Esnet		1													48						2								
FACET		1		2		10						1				148	1			3	10					1	3		
Fermilab AC		10	2	5	2			36			4	1	1	2		1	1924			2	1	45		3	4	81	3	2	2
HFIR		3	13	77				2	3	2	2	29		2						492	2	2	4		3		235	7	
JGI			2	3								1	39				2	2	957	1	8					1	2		
LCLS		4	65	77				1	1		4						3	1	2	1	829	14	3		1		1	93	
NERSC	40	281	90	26	71	4	3	44	19	6	16	16	86	64	2	10	45	4	8	14	6332	2	41	304	142	13	14	30	
NSLS II			5	25					12	1	2									3	2	95				1	6	1	
NSTX-U	35	2		1									87				3				41		356	4					
OLCF	2	169		1	15		1	14	2	4		3	14	13			4	3		1	304		4	1107	1	4			
RHIC		2	1			2		29			1						81				142			1	1015	1			
SNS		4	15	163				7	3	2		74		3		1	3	235	1	1	13	1		4	1	843	7		
SSRL		3	223	183	3			4	3	8	8	1		12			3	2	7	2	93	14	6			7	1626	31	
TMF		2	96	5				1	2	2	1	2	1	3			2				30	1					31	677	

<b>Total</b>																													
<b>Crossover</b>																													
<b>Users</b>	128	556	867	1123	138	17	24	186	80	57	231	176	240	216	3	32	207	386	61	271	1395	58	173	559	260	538	613	179	
<b>Crossover %</b>	57%	56%	34%	21%	12%	23%	6%	12%	16%	11%	44%	31%	43%	30%	6%	22%	11%	78%	6%	33%	22%	61%	49%	50%	26%	64%	38%	26%	



# Office of Science and Big Data Growth

[http://science.energy.gov/~media/ascr/pdf/programdocuments/docs/ascr-eod-workshop-2015-report\\_160524.pdf](http://science.energy.gov/~media/ascr/pdf/programdocuments/docs/ascr-eod-workshop-2015-report_160524.pdf)

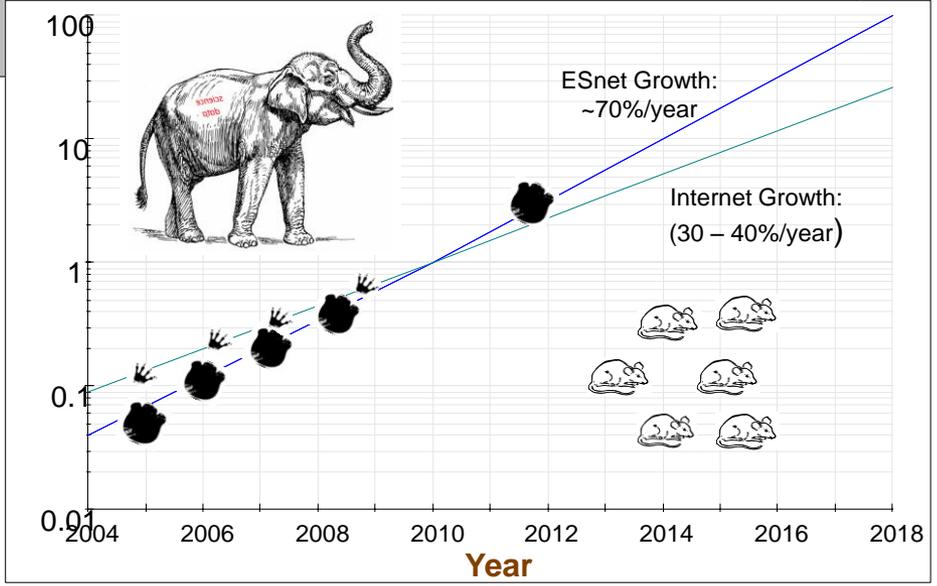
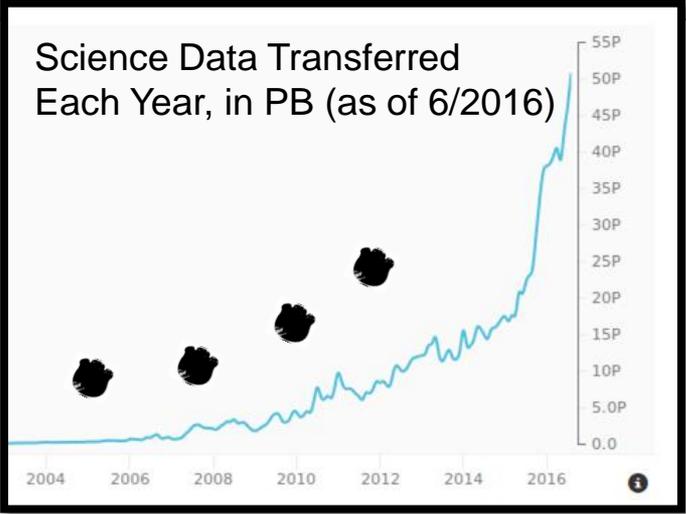
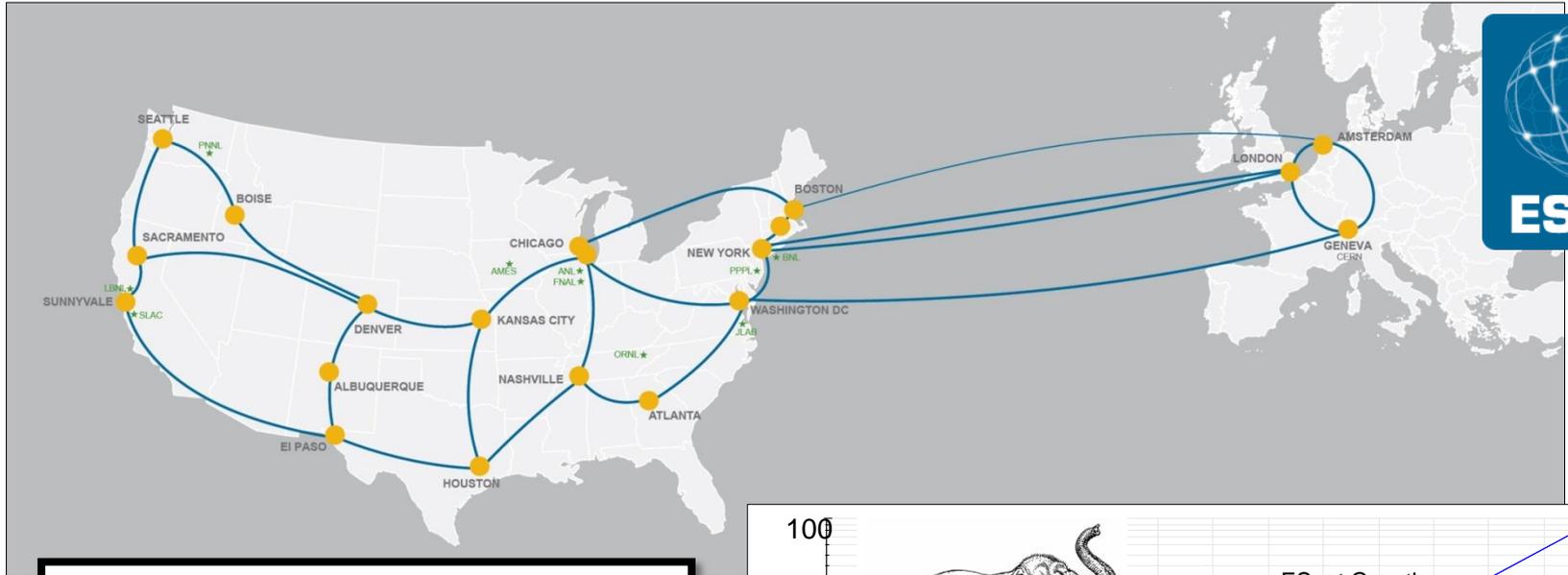
	Current/ Near Term		Long Term	
Use Case	Max rate	Monthly/Annual Totals	Max Rate	Monthly/Annual Totals
EMSL	100 Mbps	3.6 TB monthly	1 GB/s	18 TB monthly
Climate Simulation	5 TB/day	1.5 PB annually	250 TB/day	50 PB annually
ARM		18 TB monthly		5 PB obs data & 1 PB model data annually
ALS	10 Gbps	140 TB monthly	80 Gbps	1.5 PB monthly
LCLS	1-10 Gbps	1.5 PB annually	100 Gbps	15 PB annually
ORNL Neutron sources	500 MB/s	.3 PB annually	5 GB/s	1 PB annually
Scanning probe and electron microscopies	10-100 GB/day		10 Mb/s	
APS	5 TB/day	2 PB annually	1 PB/day	500 – 1000 PB annually
DUNE	10 GB/s	500-800 TB annually	100 GB/s	10-20 PB annually

# Big Data, Big Compute, Big Networks

---

- “... trend toward the convergence of data and computing.”  
[Management, Analysis, and Visualization of Experimental and Observational Data: The Convergence of Data and Computing](#)
- “... it would be a mistake to think of them [“big data” and “big compute”] as independent activities. Instead, their requirements are tightly intertwined since they both contribute to a shared goal of scientific discovery.”  
[ASCAC report on Synergistic Challenges in Data-Intensive Science and Exascale Computing](#)
- For many problems there is a deep coupling of observation (measurement) and computation (simulation)

# Data-Intensive Science Drives Exponential Network Growth

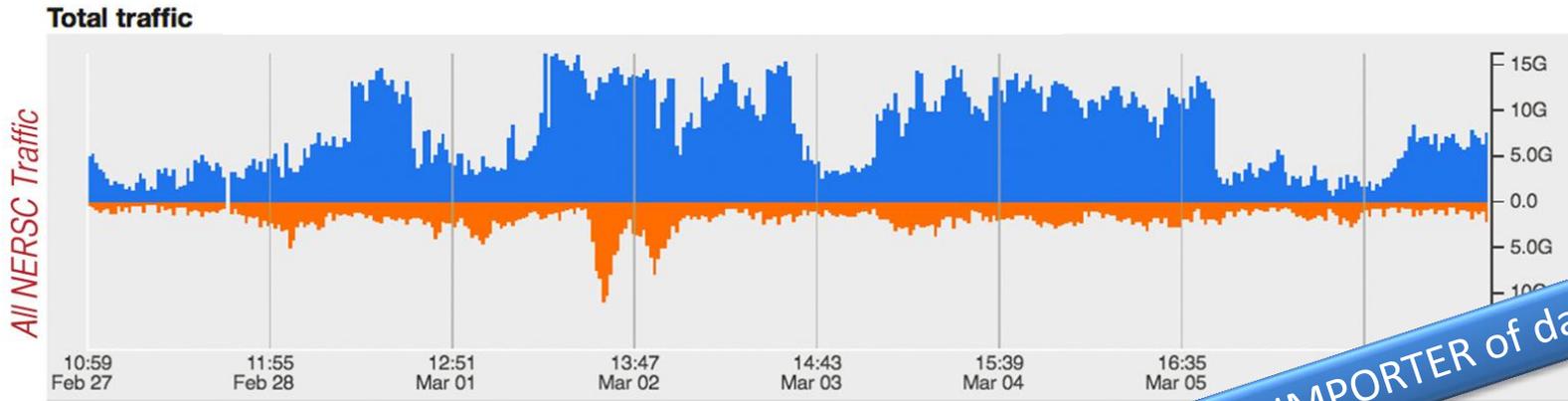


# Single Light Source Data Flow Triples NERSC's Network Usage

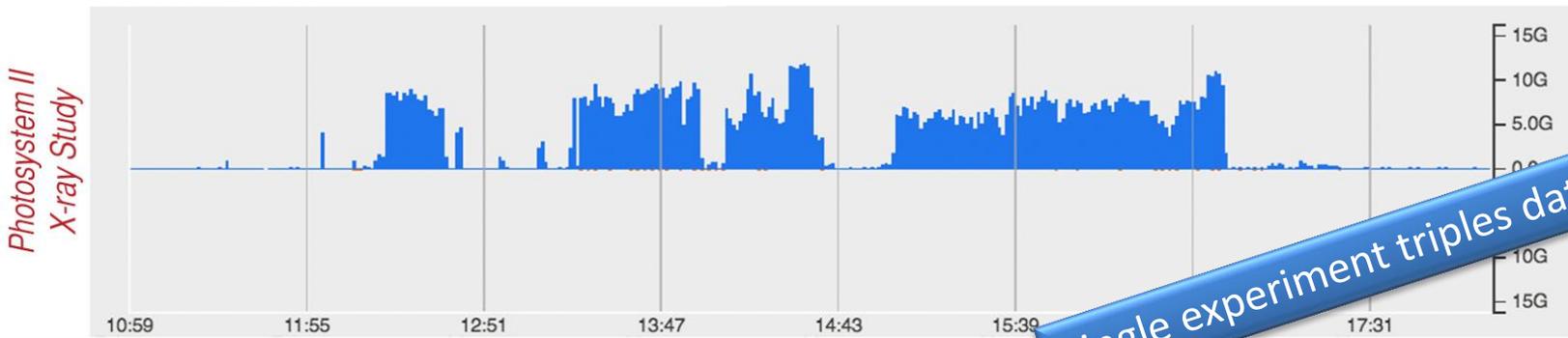


From : Wed Feb 27 10:59:00 2013 To : Thu Mar 7 10:59:00 2013

■ To site ■ From site



**Net IMPORTER of data**

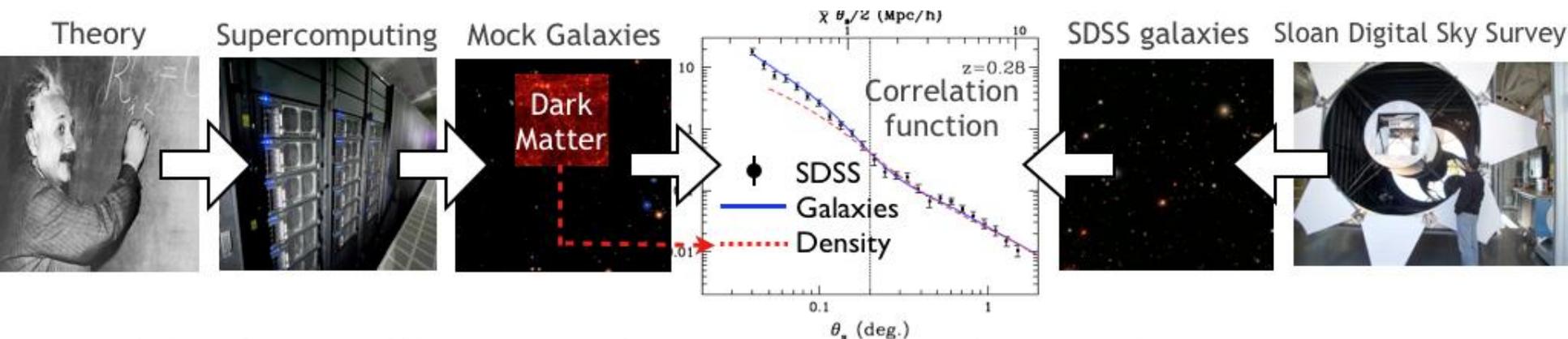


**Single experiment triples data rate**

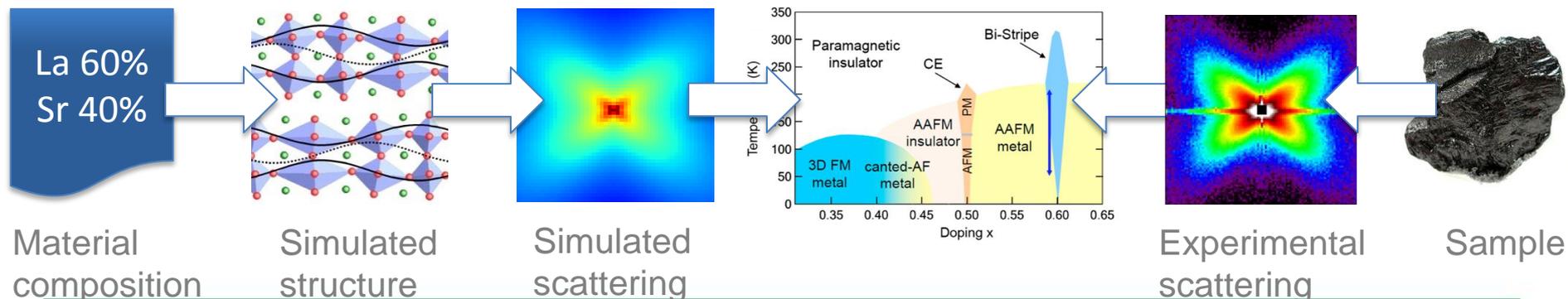
# Data-Driven Science Examples

For many problems there is a deep coupling of observation (measurement) and computation (simulation)

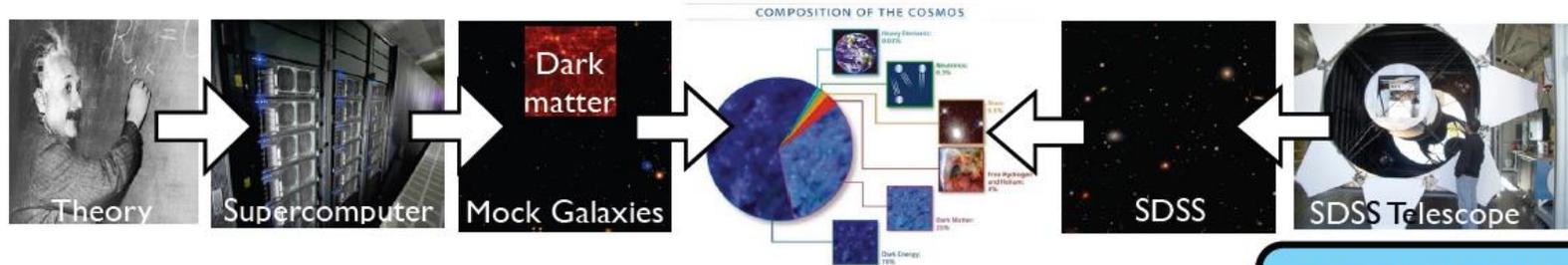
Example: Cosmology – the study of the universe as a dynamical system



Materials science: Diffuse scattering to understand disordered structures



# Computational Cosmology: Role of Simulations

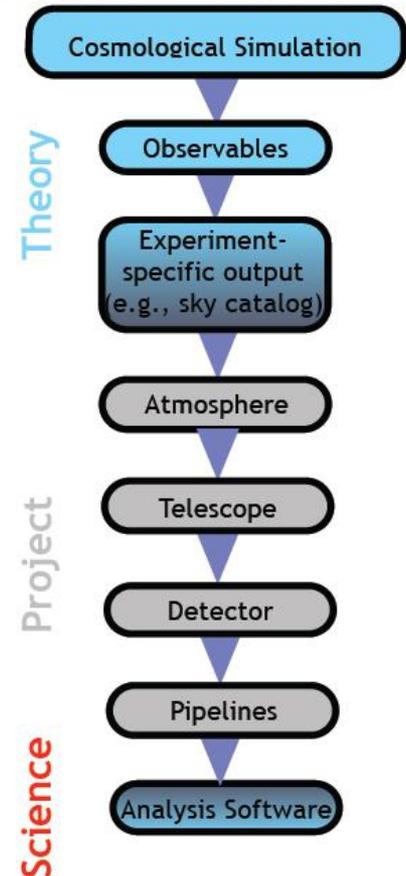


- **Three Roles of Cosmological Simulations**

- Basic theory of cosmological probes
- Production of high-fidelity ‘mock skies’ for end-to-end tests of the observation/analysis chain
- Essential component of analysis toolkits: Control systematics

- **Extreme Simulation and Analysis Challenges**

- Large dynamic range simulations; control of subgrid modeling and feedback mechanisms
- Design and implementation of **complex analyses** on large datasets; new fast (approximate) algorithms
- Solution of large statistical **inverse problems** of scientific inference (many parameters, ~10-100) at the **~1% level**



# Data Driven Neutron Scientific Discovery Enabled by HPC

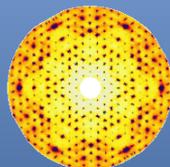
## Data Base Generation

### Neutron modeling

DISCUS  
SIMULATION PACKAGE

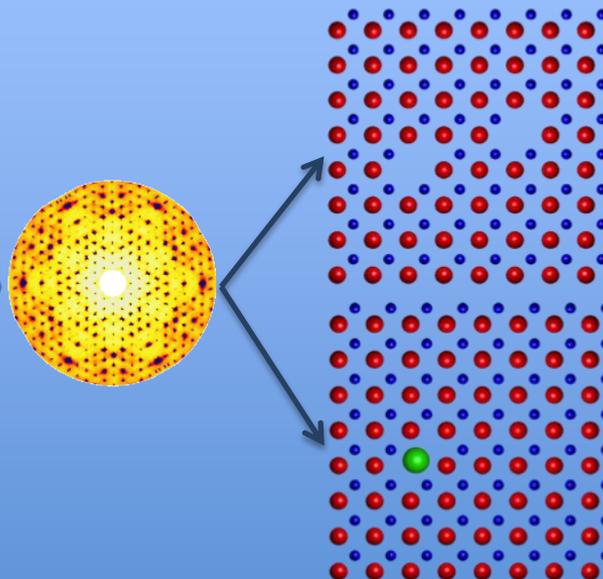


### Experimental Neutron Data



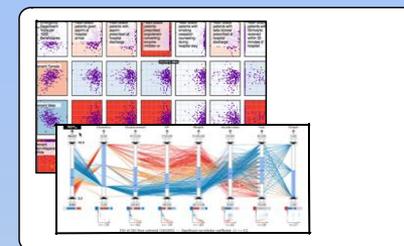
Utilize both computational and experimental production of diffuse scattering data

## Analysis and Learning



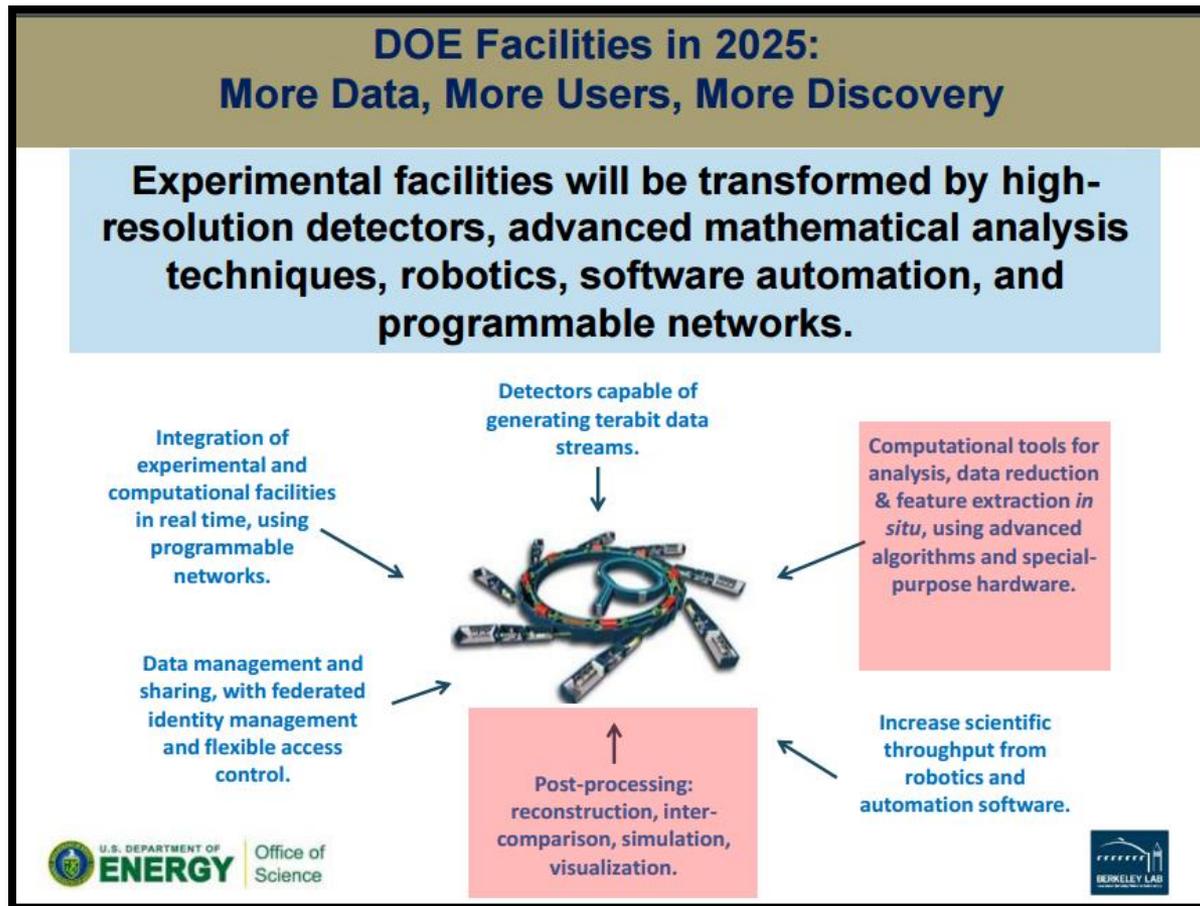
Computational methods can be generated to identify material properties and anomalies from massive & expanding database

## Scientific Discovery



Scientific analysis and search tools providing access to massive database of diffuse scattering knowledge

# More Data, More Users, More Discovery



*Applied Mathematics for Experimental Science*, J.Sethian, ASCAC July, 2015  
[http://science.energy.gov/~media/ascr/ascac/pdf/meetings/20150727/Sethian\\_ascac\\_july\\_2015.pdf](http://science.energy.gov/~media/ascr/ascac/pdf/meetings/20150727/Sethian_ascac_july_2015.pdf)



# Reports

## **Synergistic Challenges in Data-Intensive Science and Exascale Computing (ASCAC Report)**

This new report discusses the natural synergies among the challenges facing data-intensive science and exascale computing, including the need for a new scientific workflow.

[http://science.energy.gov/~media/ascr/asac/pdf/reports/2013/ASCAC\\_Data\\_Intensive\\_Computing\\_report\\_final.pdf](http://science.energy.gov/~media/ascr/asac/pdf/reports/2013/ASCAC_Data_Intensive_Computing_report_final.pdf)

## **Data Crosscutting Requirements Review**

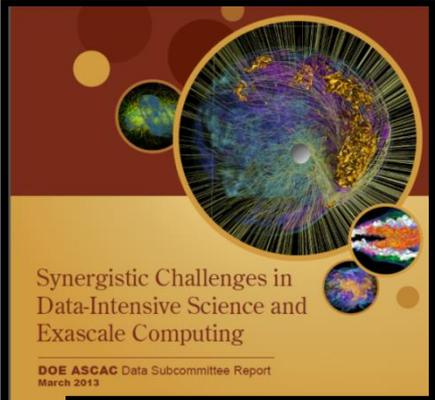
In April 2013, a diverse group of researchers from the U.S. Department of Energy (DOE) scientific community assembled in Germantown, Maryland to assess data requirements associated with DOE-sponsored scientific facilities and large-scale experiments.

[http://science.energy.gov/~media/ascr/pdf/program-documents/docs/ASCR\\_DataCrosscutting2\\_8\\_28\\_13.pdf](http://science.energy.gov/~media/ascr/pdf/program-documents/docs/ASCR_DataCrosscutting2_8_28_13.pdf)

## **Management, Analysis, and Visualization of Experimental and Observational Data: The Convergence of Data and Computing**

The purpose of this workshop...is to help the Advanced Scientific Computing Research (ASCR) and research community better understand needs related to the management, analysis, and visualization of experimental and observational data (EOD) collected and generated by experimental and observational science projects (EOS) at Office of Science user facilities.

[http://science.energy.gov/~media/ascr/pdf/programdocuments/docs/ascr-eod-workshop-2015-report\\_160524.pdf](http://science.energy.gov/~media/ascr/pdf/programdocuments/docs/ascr-eod-workshop-2015-report_160524.pdf)



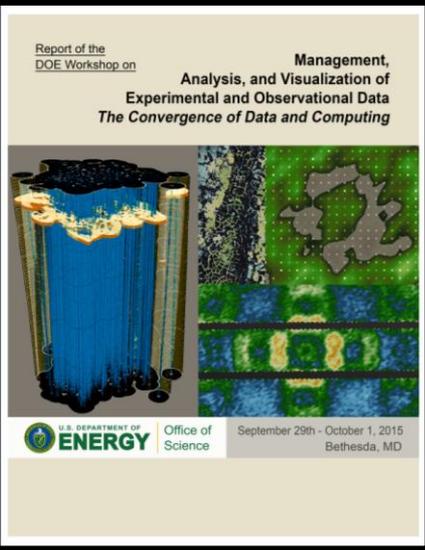
### Synergistic Challenges in Data-Intensive Science and Exascale Computing

DOE ASCAC Data Subcommittee Report  
March 2013

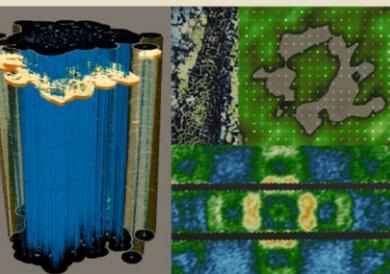


### Data Crosscutting Requirements Review

April 4-5, 2013



### Report of the DOE Workshop on Management, Analysis, and Visualization of Experimental and Observational Data The Convergence of Data and Computing



U.S. DEPARTMENT OF ENERGY

Office of Science

September 29th - October 1, 2015  
Bethesda, MD

## Selected Findings and Excerpts:

- All Experimental and Observational Science (EOS) projects represented at this workshop struggle to keep up with the demands and opportunities that a flood of data offers.
- Scientific data is increasingly at risk of being unusable.
- EOS projects are impeded due to significant “data lifecycle” needs that are largely unmet.
- “. . . our only archival process right now is that provided by the published journal. ” (Neutron source case study)



# Perspectives on Data (from Washington)

## OECD Principles and Guidelines for Access to Research Data from Public Funding



### G8 Open Data Charter

Preamble

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

February 22, 2013

MEMORANDUM FOR THE HEADS OF

FROM: John P. Holdren *JPH*  
Director

SUBJECT: Increasing Access to the Res

### 1. Policy Principles

The Administration is committed to ensuring that, to the greatest extent possible and consistent with law, the results of federally funded scientific research are made available to the scientific community. Such results include:

Scientific research supported by the Federal Government that drive our economy. The results of that research are made available for progress in areas such as health, energy, and the environment.

Access to digital data sets resulting from federal research resources and efforts on understanding and managing data underpins the forecasting industry, and

DEPARTMENT OF ENERGY

## Final Report Implementing Office of Management and Budget Information Dissemination Quality Guidelines

BY: Office of the Chief Information Officer, Department of Energy (DOE).

ON: Notice.

ARY: DOE gives notice of the final report to the Office of Management and

(OMB) that contains final

and maximize the quality, t

disseminates to members of

government wide guidelin

Appropriations Act f

guidelines in the fin



Search SC Website SC Site Search

Programs Laboratories User Facilities Universities Funding Opportunities Discovery & Innovation News

You are here: SC Home > Funding Opportunities > Statement on Digital Data Management

### Funding Opportunities

Grants & Contracts Support

Award Search

Find Funding

Statement on Digital Data Management

Suggested Elements for a Data Management Plan (DMP)

Frequently Asked Questions

Data Management

### Statement on Digital Data Management

Print Text Size: A A A Feedback: 1 | Share

The Office of Science mission is to deliver the scientific discoveries and major scientific tools that transform our understanding of nature and advance the energy, economic, and national security of the United States. The Office of Science Statement on Digital Data Management has been developed with input from a variety of stakeholders in this mission.

Here, data management involves all stages of the digital data life cycle including capture, analysis, sharing, and preservation. The focus of this statement is sharing and preservation of digital research data. To the greatest extent and with the constraints possible, data sharing should make digital research data available to and useful for the public, industry, and scientific community.

Table of Contents

Principles



THE DIRECTOR

EXECUTIVE OFFICE OF THE PRESIDENT  
OFFICE OF MANAGEMENT AND BUDGET INFORMATION  
WASHINGTON, D.C. 20503

May 9, 2013

M-13-13

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Sylvia M. Burwell *SMBurwell*  
Director

Steven VanRoekel *SVanRoekel*  
Federal Chief Information Officer

Todd Park *Toddy*  
U.S. Chief Technology Officer

Dominic J. Mancini *DominicMancini*  
Acting Administrator, Office of Information Management



## U.S. OPEN DATA ACTION PLAN

May 9, 2014



# Policy and Implementation

EXECUTIVE OFFICE OF THE PRESIDENT  
OFFICE OF SCIENCE AND TECHNOLOGY POLICY  
WASHINGTON, D.C. 20503

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren  
Director

SUBJECT: Increasing Access to the Results of Federally Funded Research and Development

## I. Policy Principles

The Administration is committed to ensuring that, to the greatest extent possible and consistent with law and the objectives of the federal government, the results of federally funded scientific research are made available to and accessible by the scientific community. Such results include peer-reviewed journal articles, reports, and other scientific publications.

Scientific research supported by the Federal Government catalyzes economic growth and innovation. The results of that research become the driving force for progress in areas such as health, energy, the environment, and national security.

Access to digital data sets resulting from federally funded research and efforts on understanding and exploiting discovered data underpins the forecasting industry, and making genome sequences available to the public has spawned many biotechnology innovations. In addition, wider access to scientific data in digital formats will create new services related to curation, preservation, analysis, and visualization. Publications and data for re-use through preservation and broad access will increase the impact and accountability of the Federal research investment. Scientific breakthroughs and innovation, promote entrepreneurship and job creation.

The Administration also recognizes that publishers provide valuable coordination of peer review, that are essential for ensuring the quality of scholarly publications. It is critical that these services continue to be supported and that Federal policy not adversely affect opportunities for research funded by the Federal Government to disseminate any analysis.

To achieve the Administration's commitment to increase access to research and digital scientific data, Federal agencies investing in research and development have clear and coordinated policies for increasing such access.

[https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)

<http://www.energy.gov/downloads/doe-public-access-plan>

<http://www.energy.gov/datanagement/doe-policy-digital-research-data-management>

<http://science.energy.gov/funding-opportunities/digital-data-management/>

## Public Access Plan



U.S. Department of Energy  
July 24, 2014

ENERGY.GOV

ENERGY.GOV

PUBLIC SERVICES SCIENCE & INNOVATION ENERGY SAVER ABOUT ENERGY.GOV OFFICES

## DOE POLICY FOR DIGITAL RESEARCH DATA MANAGEMENT

### DOE Policy for Digital Research Data Management

Resources at DOE  
Suggested Elements for a Data Management Plan  
FAQs

Glossary

The Department of Energy (DOE) is responsible for scientific and technological innovation in support of environmental cleanup of the nation's nuclear waste.

This policy is part of the implementation of the Department's research mission.

Here, data management involves all stages of the data life cycle including capture, analysis, sharing, and preservation.

This policy applies to Unclassified and Otherwise Restricted federal employees, National Laboratory and other federal grantees, and other contractor entities where law, regulation, agreement terms and conditions, or other requirements apply.

### TABLE OF CONTENTS

- Principles
- Roles and Responsibilities
  - DOE Sponsoring Research Offices
  - Respondents to DOE Research Funding
  - Recipients of DOE Research Funding
- Requirements and Guidance from DOE Sponsor
  - Requirements
  - Guidance (including Suggested Elements for a Data Management Plan)
  - Information about Data Management Resources

SC Home Organization Jobs Contact DOE Home

U.S. DEPARTMENT OF ENERGY Office of Science

Search SC Website SC Site Search GO

Programs Laboratories User Facilities Universities Funding Opportunities Discovery & Innovation News About

You are here: SC Home > Funding Opportunities > Statement on Digital Data Management

## Funding Opportunities

Grants & Contracts Support

Award Search

Find Funding

Early Career Research Program

Statement on Digital Data Management

Suggested Elements for a Data Management Plan

Frequently Asked Questions

Resources at the Office of Science User Facilities

Acknowledgements of Federal Support

## Statement on Digital Data Management

Print Text Size: A A A Feedback Share Page

The Office of Science mission is to deliver the scientific discoveries and major scientific tools that transform our understanding of nature and advance the energy, economic, and national security of the United States. The Office of Science Statement on Digital Data Management has been developed with input from a variety of stakeholders in this mission<sup>1</sup>.

Here, data management involves all stages of the digital data life cycle including capture, analysis, sharing, and preservation. The focus of this statement is sharing and preservation of digital research data.

### Table of Contents

- Principles
- Requirements
- Additional Guidance (including suggested elements for Data Management Plan)
- Additional Requirements and Guidance from Office of Science Program Offices
- Information about Data Management Resources at Office of Science User Facilities
- Glossary
- FAQs
- References

DEPARTMENT OF ENERGY  
PAGES  
Public Access Gateway  
for Energy & Science

Introduction to DOE PAGES DOE PAGES Advanced Search

### What does "Beta" mean?

DOE PAGES contains an initial collection of journal articles and accepted manuscripts as a demonstration of its functionality and eventual expanded content. Over the next year, additional metadata and links to articles and accepted manuscripts will be added as they are submitted to OSTI, with anticipated annual growth of 20,000-30,000 publicly-accessible articles and manuscripts. When DOE PAGES moves beyond the "beta" period, it will offer distributed full-text access to all DOE-affiliated accepted manuscripts or articles after an administrative interval of 12 months.

### Guidance for DOE-Funded Authors

I'm a researcher at a DOE national laboratory and have just had a manuscript accepted for publication in a peer-reviewed journal; what do I need to do in order to comply with DOE's public access requirements?

I'm a researcher with a grant from DOE and have just had a manuscript accepted for publication in a peer-reviewed journal; what do I need to do in order to comply with DOE's public access requirements?

### Find out more

Do you have questions about DOE PAGES content, procedures, or policies? More information is available at OSTI's Public Access Policy page and in our Frequently Asked Questions.

<http://www.osti.gov/pages/>

# Brief History

---

- **COMPETES 2010 “Interagency Public Access Committee”**
- Office of Science Working Group on Digital Data
- Office of Science FACA Reports (2011)
- OSTP Request for Information (2012)
- Office of Science User Facility Input (2013)
- **OSTP Memo “Increasing Access to the Results of Federally Funded Scientific Research” (Feb., 2013)**
- **DOE Public Access Plan and Office of Science Statement on Digital Data Management (July, 2014)**
- **DOE Policy for Digital Research Data Management (Sept, 2015)**

# Office of Science Statement on Digital Data Management

<http://science.energy.gov/funding-opportunities/digital-data-management/>

[SC Home](#) [Organization](#) [Jobs](#) [Contact](#)

[DOE Home](#) »



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

Search SC Website

SC Site Search



GO

[Programs](#)

[Laboratories](#)

[User Facilities](#)

[Universities](#)

[Funding Opportunities](#)

[Discovery & Innovation](#)

[News](#)

[About](#)

You are here: [SC Home](#) » [Funding Opportunities](#) » [Statement on Digital Data Management](#)

## Funding Opportunities

[Grants & Contracts Support](#)

[Award Search](#)

[Find Funding](#)

[Early Career Research Program](#)

[Statement on Digital Data Management](#)

[Suggested Elements for a Data Management Plan](#)

[Frequently Asked Questions](#)

[Resources at the Office of Science User Facilities](#)

[Acknowledgements of Federal Support](#)

## Statement on Digital Data Management

Print | Text Size: [A](#) [A](#) [A](#)

[Feedback \[+\]](#) | [Share Page](#) ▾

The Office of Science mission is to deliver the scientific discoveries and major scientific tools that transform our understanding of nature and advance the energy, economic, and national security of the United States. The Office of Science Statement on Digital Data Management has been developed with input from a variety of stakeholders in this mission<sup>1</sup>.

Here, data management involves all stages of the digital data life cycle including capture, analysis, sharing, and preservation. The focus of this statement is [sharing](#) and [preservation](#) of digital research data.

### Table of Contents

- [Principles](#)
- [Requirements](#)
- [Additional Guidance \(including suggested elements for Data Management Plan\)](#)
- [Additional Requirements and Guidance from Office of Science Program Offices](#)
- [Information about Data Management Resources at Office of Science User Facilities](#)
- [Glossary](#)
- [FAQs](#)
- [References](#)

# Office of Science Statement on Digital Data Management

---

## Principles

- Effective data management has the potential to increase the pace of scientific discovery and promote more efficient and effective use of government funding and resources. Data management planning should be an integral part of research planning.
- Sharing and preserving data are central to protecting the integrity of science by facilitating validation of results and to advancing science by broadening the value of research data to disciplines other than the originating one and to society at large. To the greatest extent and with the fewest constraints possible, and consistent with the requirements and other principles of this Statement, data sharing should make digital research data available to and useful for the scientific community, industry, and the public.
- Not all data need to be shared or preserved. The costs and benefits of doing so should be considered in data management planning.



# Office of Science Statement on Digital Data Management

---

- Requirements apply to proposals for research funding
- Requirements apply to proposals submitted for new, renewal, and some supplemental research funding
- Requirements apply to proposals regardless of the PI's institution
- Requirements do *not* apply to applications to use Office of Science user facilities.
- SC sponsored research activities at the DOE National Laboratories for which a DOE-approved DMP does not already exist will be required to develop a DMP.
- DMPs will be reviewed as part of the overall Office of Science research proposal merit review process.
- Additional requirements and review criteria for the DMP may be identified by the sponsoring program or sub-program, or in the solicitation.

# Office of Science Statement on Digital Data Management

<http://science.energy.gov/funding-opportunities/digital-data-management/>

## Requirements (abridged)

1. DMPs should describe whether and how data generated in the course of the proposed research will be shared and preserved. DMPs must describe how data sharing and preservation will enable validation of results, or how results could be validated if data are not shared or preserved.
2. DMPs should provide a plan for making all research data displayed in publications resulting from the proposed research open, machine-readable, and digitally accessible to the public at the time of publication. The published article should indicate how these data can be accessed.
3. DMPs should consult and reference available information about data management resources to be used in the course of the proposed research. In particular, DMPs that explicitly or implicitly commit data management resources at a facility beyond what is conventionally made available to approved users should be accompanied by written approval from that facility.
4. DMPs must protect confidentiality, personal privacy, Personally Identifiable Information, and U.S. national, homeland, and economic security; recognize proprietary interests, business confidential information, and intellectual property rights; avoid significant negative impact on innovation, and U.S. competitiveness; and otherwise be consistent with all applicable laws, regulations, and DOE orders and policies. There is no requirement to share proprietary data.



# Definitions

---

## Digital Research Data:

The term *digital data* encompasses a wide variety of information stored in digital form including: experimental, observational, and simulation data; codes, software and algorithms; text; numeric information; images; video; audio; and associated metadata. It also encompasses information in a variety of different forms including raw, processed, and analyzed data, published and archived data.

This statement focuses on *digital research data*, which are *research data* that can be stored digitally and accessed electronically. OMB Circular A110 defines *research data* as follows:

“Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This 'recorded' material excludes physical objects (e.g., laboratory samples). Research data also do not include:

- (A) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and
- (B) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.”

# Definitions

---

## **Data Preservation:**

*Data preservation* means providing for the usability of data beyond the lifetime of the research activity that generated them.

## **Data Sharing:**

*Data sharing* means making data available to people other than those who have generated them. Examples of data sharing range from bilateral communications with colleagues, to providing free, unrestricted access to the public through, for example, a web-based platform.

## **Validate:**

In the context of this statement, *validate* means to support, corroborate, verify, or otherwise determine the legitimacy of the research findings. Validation of research findings could be accomplished by reproducing the original experiment or analyses; comparing and contrasting the results against those of a new experiment or analyses; or by some other means.

# Additional Guidance and Requirements from Program Offices

The screenshot shows the top navigation bar with the U.S. Department of Energy Office of Science logo and search fields. Below the navigation bar, a breadcrumb trail reads: "You are here: SC Home » Programs » ASCR Home » Funding Opportunities » Additional Requirements and Guidance for Digital Data Management". The main content area features a blue header for "Advanced Scientific Computing Research (ASCR)" and a sub-header for "Funding Opportunities Additional Requirements and Guidance for Digital Data Management". A left sidebar contains a menu with "Funding Opportunities" highlighted. The bottom of the sidebar shows a link to "2013 Exascale Operating".

The screenshot shows the top navigation bar with the U.S. Department of Energy Office of Science logo and search fields. Below the navigation bar, a breadcrumb trail reads: "You are here: SC Home » Programs » NP Home » Funding Opportunities » Additional Requirements and Guidance for Digital Data Management". The main content area features a blue header for "Nuclear Physics (NP)" and a sub-header for "Funding Opportunities Additional Requirements and Guidance for Digital Data Management". A left sidebar contains a menu with "Funding Opportunities" highlighted. The main text area begins with "The Office of Nuclear Physics (NP) supports the Digital Data Principles enunciated in the Office of Science Statement on Digital Data Management. The following requirements and guidance provided by the NP program office supplement but do not replace the requirements and guidance in the Office of Science Statement on Digital Data Management." and includes a section for "Program Requirements".



 GO

- Programs
- Laboratories
- User Facilities
- Universities
- Funding Opportunities
- Discovery & Innovation
- News
- About

You are here: SC Home » Funding Opportunities » Statement on Digital Data Management » Data Management Resources at the Office of Science User Facilities

## Funding Opportunities

Grants & Contracts Support

Award Search

Find Funding

Early Career Research Program

Statement on Digital Data Management

Suggested Elements for a Data Management Plan

Frequently Asked Questions

Resources at the Office of Science User Facilities

Acknowledgements of Federal Support

**CONTACT INFORMATION**

**Office of Science**  
U.S. Department of Energy  
1000 Independence Ave., SW  
Washington, DC 20585

# Statement on Digital Data Management

## Data Management Resources at the Office of Science User Facilities

Print | Text Size: A A A

Feedback (+) | Share Page ▾

Follow the links to learn more about data management resources at each of the Office of Science User Facilities. If you do not see your facility of interest listed here, please consult the appropriate [Office of Science Program](#) page.

### Advanced Scientific Computing Research (ASCR)

Facility	Host Institution	Data Management Resources
National Energy Research Scientific Computing Center (NERSC)	LBL	<a href="#">Link</a>
Argonne Leadership Computing Facility (ALCF)	ANL	<a href="#">Link</a>
Oak Ridge Leadership Computing Facility (OLCF)	ORNL	<a href="#">Link</a>
Energy Sciences Network (ESnet)	LBL	<a href="#">Link</a>

### Basic Energy Sciences (BES)

Facility	Host Institution	Data Management Resources
<i>Light Sources</i>		
Advanced Light Source (ALS)	LBL	<a href="#">Link</a>
Advanced Photon Source (APS)	ANL	<a href="#">Link</a>
Linac Coherent Light Source (LCLS)	SLAC	<a href="#">Link</a>
National Synchrotron Light Source (NSLS)	BNL	<a href="#">Link</a>

# Suggested Elements for a Data Management Plan



## Funding Opportunities

Grants & Contracts  
Support

Award Search

Find Funding

Early Career Research  
Program

Statement on Digital Data  
Management

**Suggested Elements for a  
Data Management Plan**

Frequently Asked Questions

Resources at the Office of  
Science User Facilities

Acknowledgements of  
Federal Support

## Statement on Digital Data Management

### Suggested Elements for a Data Management Plan

Print Text Size: A A A

Feedback [+] Share Page ▾

The Principal Investigator should determine which data should be the subject of the DMP and, in the DMP, propose which data should be shared and/or preserved in accordance with the [Requirements for Digital Data Management](#).

The following list of elements for a DMP provides suggestions regarding the structure of the DMP:

- **Data Types and Sources.** A brief, high-level description of the data proposed research and which of these are considered [digital research data](#).
- **Content and Format.** A statement of plans for data and metadata, a description of documentation plans, annotation of relevant software, standards. (Existing, accepted community standards should be used where possible. Where community standards are missing or inadequate, the DMP could propose alternate strategies that facilitate sharing, and should advise the sponsoring program of any need to develop or generalize standards.)
- **Sharing and Preservation.** A description of the plans for data sharing and preservation. This should include, when appropriate:
  - the anticipated means for sharing and the rationale for any restrictions on who may access the data and under what conditions;
  - a timeline for sharing and preservation that addresses both the minimum length of time the data will be available and

- Data Types and Sources
- Content and Format
- Sharing and Preservation
- Protection
- Rationale

# Data ID Service

<http://www.osti.gov/home/doi-data-id-service>



Speeding access to science information from DOE and beyond



HOME

ABOUT OSTI

SCIENCE SEARCH  
TOOLS

DOE PAGES<sup>Beta</sup> /  
PUBLIC ACCESS

COMMUNICATIONS

DOE STI PROGRAM

## DOE Data ID Service

[DataCite](#) | [Contact DOE Data ID Service](#)

### DOE Data ID Service

The DOE Office of Scientific and Technical Information (OSTI) offers a service for registering datasets to help increase access to digital data from DOE-funded scientific research. Through the DOE Data ID Service, OSTI assigns persistent identifiers, known as Digital Object Identifiers (DOIs), to datasets submitted by DOE and its contractor and grantee researchers and registers the DOIs with DataCite to aid in citation, discovery, retrieval, and reuse.

DOIs can be used to search for datasets on a number of OSTI's scientific and technical information database products. OSTI assigns and registers DOIs for datasets for DOE researchers as a free service to enhance the Department of Energy's management of this important resource.

OSTI, the only U.S. government member of DataCite, can assign DOIs to other federal agencies' datasets on a cost-reimbursable basis. For information about this interagency service, please [contact DOE Data ID Service](#).

[DOE Data ID Service Fact Card](#) (654.74KB pdf)



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

# SC Funding Opportunity Announcement Language

## APPENDIX 6: DATA MANAGEMENT PLAN

### **Provide a Data Management Plan (DMP) that addresses the following requirements:**

1. DMPs should describe whether and how data generated in the course of the proposed research will be shared and preserved. If the plan is not to share and/or preserve certain data, then the plan must explain the basis of the decision (for example, cost/benefit considerations, other parameters of feasibility, scientific appropriateness, or limitations discussed in #4). At a minimum, DMPs must describe how data sharing and preservation will enable validation of results, or how results could be validated if data are not shared or preserved.
2. DMPs should provide a plan for making all research data displayed in publications resulting from the proposed research open, machine-readable, and digitally accessible to the public at the time of publication. This includes data that are displayed in charts, figures, images, etc. In addition, the underlying digital research data used to generate the displayed data should be made as accessible as possible to the public in accordance with the principles stated in the Office of Science Statement on Digital Data Management (<http://science.energy.gov/funding-opportunities/digital-data-management/>). This requirement could be met by including the data as supplementary information to the published article, or through other means. The published article should indicate how these data can be accessed.
3. DMPs should consult and reference available information about data management resources to be used in the course of the proposed research. In particular, DMPs that explicitly or implicitly commit data management resources at a facility beyond what is conventionally made available to approved users should be accompanied by written

# SC Funding Opportunity Announcement Language

## Merit Review Criteria:

### SCIENTIFIC AND/OR TECHNICAL MERIT OF THE PROPOSED RESEARCH

- What is the scientific innovation of proposed effort?
- How does the proposed work compare with other efforts in its field, both in terms of scientific and/or technical merit and originality?
- How might the results of the proposed work impact the direction, progress, and thinking in relevant scientific fields of research?
- What is the likelihood of achieving influential results?
- Is the Data Management Plan suitable for the proposed research and to what extent does it support the validation of research results?

