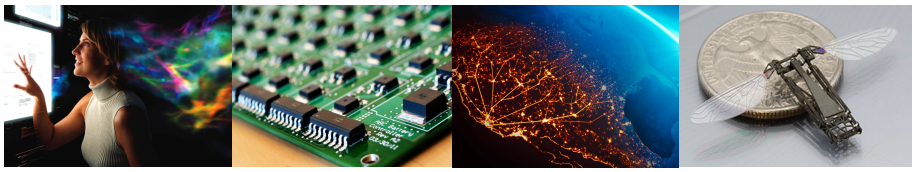
# Open Science and Public Access: an NSF Perspective



**Jim Kurose**
**Assistant Director, NSF, Computer & Information Science & Engineering**
**Co-Chair, US NSTC CoS Interagency Working Group on Open Science**

**The National Academies: Toward an Open Science Enterprise**
July 20, 2017

## Overview

- NSF policies and activities in public access
- Five (personal) observations

# NSF Public Access

- Clear and open communication of research results is central to the progress of science
  - publications (traditional)
  - data (newer)

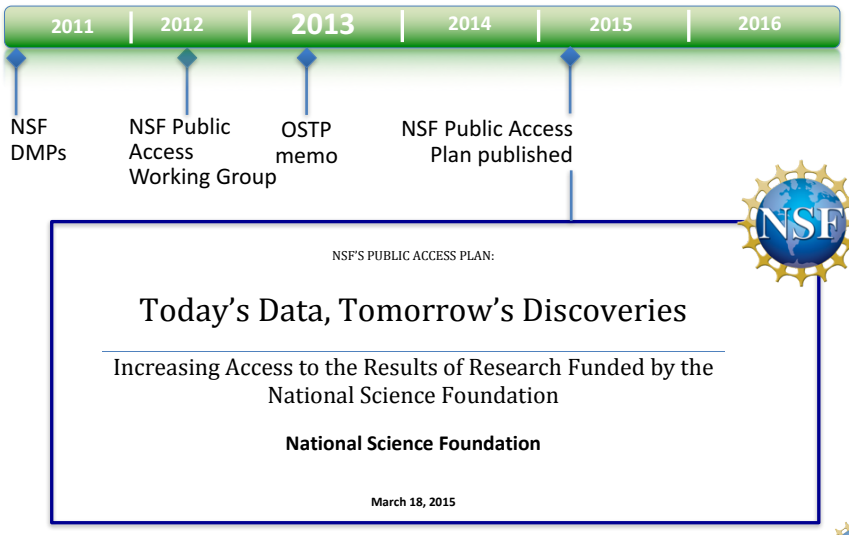
[NSF] "continues its commitment to expand public access to the results of its funded research. Public access is intended to accelerate the dissemination of fundamental research results that will advance the frontiers of knowledge and help ensure the nation's future prosperity"

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

# NSF Public Access: a brief history


**US Office of Science and Technology Policy (OSTP):** directs 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: a brief history




# NSF Public Access: a brief history


NSF Public Access Repository online (par.nsf.gov)

- Provide public access to journal, juried conference papers
- PI's must deposit publications in PAR, awards made FY 2016 onwards
- Collaboration with US Dept. of Energy (PAGES)

## Public Access: recent USG activities (1/17)


### US Government-wide (Jan. 2017)

- All subject Federal agencies have posted public access plans (compliance with 2013 OSTP memo)
- Published: "Principles for Promoting Access to Federal Government-Supported Scientific Data and Research Findings Through International Scientific Cooperation"
- New Open Science Interagency Working Group formed (NIH, NSF co-chairs)

## NSF & Data Management

- Proposal & Award Policies & Procedures Guide (PAPPG): XI D4:

  "Investigators are expected to share with other researchers, at no more than incremental cost and 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." … "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."

- NSF Public Access Plan (NSF 15-52, section 3.2):

  "To the extent feasible and consistent with applicable law and policy; agency mission; resource constraints; U.S. national, homeland, and economic security, digitally formatted scientific data resulting from unclassified research supported wholly or in part by NSF funding should be stored and publicly accessible to search, retrieve, and analyze. NSF requires applicants for funding to prepare a DMP. The current requirement, which was a January 18, 2011 implementation of the Foundation's long-standing data-sharing policy … "

# NSF Public Access Plan: data

## Data Management Plans (DMP):

- "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]
- Bottom-up implementation, top-down guiding principles
  - *"one size" does not fit all* of science and engineering

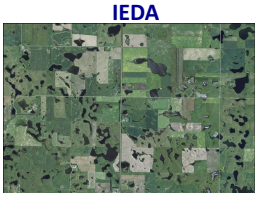
# NSF Public Access Plan: data

## Data Management Plans (DMP):

- Additional DMP guidance provided by individual directorates, divisions: https://www.nsf.gov/bfa/dias/policy/dmp.jsp
- DMP policy and practice defined:
  - Proposal & Award Policies & Procedures Guide (PAPPG) Chapter XI.D.4 (Intellectual Property: Dissemination and Sharing of Research Results)
  - NSF PA plan: https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf15052
  - Directorate, divisional DMP guidance
  - Individual solicitations

# Repositories: domain-appropriate data solutions

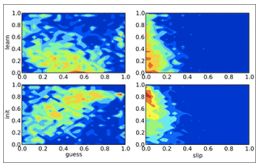
**IEDA**



**Solid Earth Data**

- Data services for observational solid earth data from the Ocean, Earth, and Polar Sciences
- Access to 15+ collections, data analysis tools
- Education, training

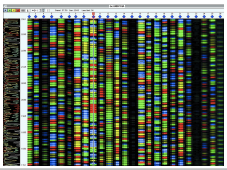
**LearnSphere**



**Educational Data**

- Maintains central metadata store of what datasets exist
- Distributed features allowing contributors control over access to their own data
- Over 30,000 research usages

**Dryad**

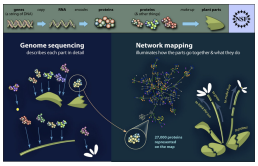


**General Purpose Digital Repository**

- Curated general-purpose repository making data underlying scientific publications discoverable, freely reusable, citable
- 2016: 15,000+ data sets, 22% increase in number of journals

# Repositories: domain-appropriate data solutions

**TAIR**




**Genetic and molecular biology data**

- Maintains a database of genetic and molecular biology data for the model higher plant Arabidopsis thaliana
- Provides centralized data access to over 30,000 Arabidopsis genes

**ICPSR**

**Social and behavioral research**

- Provides leadership and training in data access, curation, and methods of analysis for the social science research community
- Includes almost 10,000 studies and a data archive of over 250,000 files

**Zooniverse**

**People-powered research**

- World's largest platform for research by volunteers
- over 1.5 million registered volunteers worldwide and over 50 active projects across the disciplines

## Overview

- NSF policies and activities in public access
- Five (personal) observations

## 1. Heterogeneity among Federal agencies

Heterogeneity in:

- agency mission
- intramural versus extramural research
  - "ownership" of data and publications
- different types of data with different considerations (e.g., privacy), and value (e.g., weather) beyond science

One size does not fit all (even within a directorate within an agency)

## 2. DMPs: bottom-up, domain-specific

- Research community defines, evaluates scientific value-proposition of DMP in scientific context
- Research community (implicitly, explicitly) defines business models for data
  - merit review, DMP guidance, repositories
- Community-driven standards (e.g., metadata)
- DMPs: part of routine, normal, everyday business of PI community, agency
  - "conversation" has changed

## 3. Importance of sustainable business models

- Realistic assumptions about funding, budget arithmetic
- Limited number of models for sustainable research data repositories:
  - structural funding (e.g., by agencies, balancing funding for repositories, research infrastructure, research itself).
  - host institutional funding (e.g., university)
  - deposit or access fees (may intersect with agency funding, policy)

## 4. Integrating PA policies into agency processes

- Agency perspective: integration into routine, existing agency business processes (e.g., proposal submission, annual report, award management)
  - "compliance by design" (e.g., publication submission into PAR, DMP submission)
  - minimizing burden (on both sides)
- Importance of engaging all stakeholders

## 5. Future Challenges

- "Additional" research products (software in particular)
- Making open content more accessible: search/discovery, permanent identifiers, machine-readable formats
- International

## Conclusion

- NSF continues its deep commitment to expand public access to results of its funded research
- Realizing the enormous potential of data requires a long-term, bold, sustainable, and comprehensive approach, by NSF and by many partners

*We are just beginning.*
*We know it must involve the entire community.*
*And we know it will take time.*

## Thanks to those who got us here!

- NSF Public Access WG leadership: (2011-now): M. Cragin, A. Friedlander, C. Gabriel, M. Gutman, P. Knezek, J. Kurose, A. Northcutt, A. Tomkins, J. Tornow
- NSF Public Access WG members (2013-now): J. Silverthorne, D. Aronson, M. Cavanaugh, J. Feldman, N. Kaplan, A. Kelly, M. Koszalka, A. Lewis, B. Mihaila, B. Miller, A. Northcutt, I. Qualters, M. Robbs, J. Silverthorne, E. Rissi, C. Cooper, A. Medina-Borja, M. Santonastasso, C. Schrein, H. Wactlar, R. Wakimoto
- Technical Teams in NSF/DIS, DOE/OSTI, NSF/BFA, NSF/SBE (25 individuals total)
- ~50 PDs and staff

# Credits

- Copyrighted material used under Fair Use. If you are the copyright holder and believe your material has been used unfairly, or if you have any suggestions, feedback, or support, please contact: ciseitsupport@nsf.gov.
- Except where otherwise indicated, permission is granted to copy, distribute, and/or modify all images in this document under the terms of the GNU Free Documentation license, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation license" at http://commons.wikimedia.org/wiki/Commons:GNU_Free_Documentation_License.
- The inclusion of a logo does not express or imply the endorsement by NSF of the entities' products, services, or enterprises.