Data citation motivation and policies

Dr. Timothy Killeen, Assistant Director, Directorate for Geosciences (GEO)

National Science Foundation
Setting the stage over the last decade

CONTEXT - ALL SCIENCE
“The capacity of this technology has crossed thresholds that now make possible a comprehensive “cyberinfrastructure” on which to build new types of scientific and engineering knowledge environments and organizations and to pursue research in new ways and with increased efficacy”

September 2003
“Long-lived digital data collections are powerful catalysts for progress and for democratization of science and education. Proper stewardship of research requires effective policy to maximize their potential”

National Science Board
September 2005
“Developing, supporting, and promoting educational efforts to effect change in the research enterprise regarding the importance of the stewardship of digital data produced by all scientific and engineering disciplines/domains.”

September 2006
“Science and engineering digital data are routinely deposited in a well-documented form, are regularly and easily consulted and analyzed by specialists and nonspecialists alike, are openly accessible while suitably protected and are reliably preserved.”

March 2007
Key characteristics of the current digital data landscape

- the products of science and the starting point for new research are increasingly digital and increasingly “born-digital”;  
- exploding volumes and rising demand for data use are driven by the rapid pace of digital technology innovations;  
- all sectors of society are stakeholders in digital preservation and access; and  
- a comprehensive framework for cooperation and coordination to manage the risks to preservation of digital data is missing.

January 2009
Existing NSF Policy – long standing

Existing policy states NSF’s expectations with respect to sharing of data and other research products:

- **Grant Conditions:** “NSF expects investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work. It also encourages grantees to share software and inventions or otherwise act to make the innovations they embody widely useful and usable.”

- **Grant Proposal Guide:** Current proposal preparation requirements stipulate that the project description should provide: “...a clear description of experimental methods and procedures and plans for preservation, documentation, and sharing of data, samples, physical collections, curriculum materials and other related research and education products.”
- Geosciences is now predominately “born-digital”
- Three examples of activities that will benefit from data citation
Era of Observation and Simulation
Science and Society Transformed by Data

- Modern geoscience
  - Data- and compute-intensive
  - Integrative, multi-scale
- Multi-disciplinary collaborations to address complexity
  - Individuals, groups, teams, communities
- Sea of Data
  - Age of Observation
  - Distributed, central repositories, sensor-driven, diverse, etc.

Data at rest & data in motion
Science, Engineering and Education for Sustainability (SEES)
Science, Engineering and Education for Sustainability

NSF's SEES portfolio supports interdisciplinary research and education.

- SEES established in Fiscal Year 2010
- Cross-directorate NSF investment
- Portfolio of existing, new and upcoming programs
- Encourage systems-based approaches
- Highlights NSF’s unique role
SEES Mission and goals

**Mission:**
To advance science, engineering, and education to inform the societal actions needed for environmental and economic sustainability and sustainable human well-being.

**Goals:**

1. *Building the knowledge base.*
   Support interdisciplinary research and education that can facilitate the move towards global sustainability.

2. *Growing the workforce of the future.*
   Develop a workforce trained in the interdisciplinary scholarship needed to understand and address the complex issues of sustainability.

3. *Forging critical partnerships.*
   Build linkages among existing projects and partners and add new participants in the sustainability research enterprise.
SEES Portfolio Framework

**Topical themes:**
- Natural Systems
- Human Systems
- Built Systems
- Energy & Materials
- Adaptation & Resilience

**Functional characteristics:**
- Systems-Thinking
- Partnerships & Networks
- Education & Learning
- Workforce Development
- Information Systems
SEES portfolio

- Ocean Acidification (OA)
- Climate Change Education Partnership (CCEP)
- Decadal and Regional Climate Prediction using Earth System Models (EaSM)
- Dimensions of Biodiversity
- Water Sustainability and Climate (WSC)
- Research Coordination Networks - SEES track (RCN-SEES)
- Dynamics of Coupled Natural and Human Systems - SEES track (CNH-SEES)
- SEES Fellows

- Sustainability Research Networks (SRN)
- Sustainable Energy Pathways (SEP)
- SEES focus in Partnerships for International Research and Education (PIRE)
- Sustainable Chemistry, Engineering and Materials (“SusChEM”)
- Coastal regions (“CoastalSEES”)
- Arctic regions (“ArcticSEES”)
- Hazards and Resilience (“HazardsSEES”)
- Information Science and Engineering (“RISES”)
Maintainability, sustainability, and extensibility
In draft form:

Cyber Infrastructure Framework for the 21st Century:

A Vision and Strategy for Data in Science and Engineering
• An alternative approach to respond to daunting science and CI challenges
• EarthCube is an outcome and a process
• EarthCube will require broad community participation

EARTH CUBE V I S I O N
Goal of Earth Cube

to transform the conduct of research in geosciences by supporting the development of community-guided cyberinfrastructure to integrate data and information for knowledge management across the Geosciences.
Long-term Goals for EarthCube

- Transform practices within the geosciences community spanning over the next decade, including raising to 50% the rate of data citation
- Provide unprecedented new capabilities to researchers and educators
- Vastly improve the productivity of community
- Accelerating research Earth system research
- Provide a uniform knowledge management framework for all of the geosciences
- More than 700 registrants to social media site
- More than 100 white papers with designs, requirements, use case examples, etc.
- More than 60 EAGER grant applications
- 9 EAGER awards have been made for a total of $1.3m
- Next Charrette: June 2012
- Community designed, owned, governed
- No winners or losers
- An NSF “experiment”; Full partnership with USGS
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp</td>
<td>66hg</td>
<td></td>
</tr>
<tr>
<td>60%</td>
<td></td>
<td>h0%</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9M 9p</td>
<td>Cr 17%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60%</td>
<td>h0%</td>
</tr>
</tbody>
</table>

◆ Be prepared | 6 iz | z ac $ 30% | 1 b | 2 1 p | G o i z H o | 2 1 p |
Education, Outreach and Diversity

$1.2B/yr
Why NSF?
Grand Challenges in Sustainability Education: Driving Questions for NSF

- How does the nation train and develop its workforce?
- How should we teach and learn science in the 21st Century?
- How does the nation create a science literate citizenry?
- How can we broaden and deepen participation in science and engineering?
- How does NSF most effectively deploy its resources in STEM education and learning?
Earth to America – a Dose of Reality

• 6th in global innovation-based competitiveness
• 11th among industrialized nations in the fraction of 25- to 34-year-olds who have graduated from high school
• 16th in college completion rate
• 22nd in broadband Internet access
• 24th in life expectancy at birth
• 27th among developed nations in the proportion of college students receiving degrees in science or engineering
• 48th in quality of K-12 math and science education

Charles Vest, NAE
Sources: National Center for Education Statistics, IPEDS Completions and Fall enrollment surveys; Higher Education Research Institute, American Freshman Survey; and U.S. Census Bureau, Current Population Survey.
Expeditions in Education (E²)

• New cross-directorate, interdisciplinary effort
• Three Initial Focus Areas
  – Transforming undergraduate STEM learning
  – Learning and understanding sustainability
  – Cyber-learning, Data, and Observations for STEM Education
• Infuse cutting-edge science and engineering into the preparation of a world-class scientific workforce for 21st century
• Ensure that all NSF education and workforce investments are drawing on the latest educational theory, research, and evidence.
NSF, NSB, Organizations, and Societies

GUIDANCE
“Dear Colleague Letter”:

- Facilitating open and equal access to data and data set
- GEO encourages data citation & encourages community to leadership to establish data citation within the geosciences as the rule rather than the exception
New Data Management Plan
Requirements

- All proposals are required to include, as a supplementary doc, a data management plan of up to 2 pages.
- Plan should describe how the proposal will conform to NSF policy on dissemination and sharing of research results.
- A valid Data Management Plan may include only the statement that no detailed plan is needed, as long as a clear justification is provided.
- Plan will be reviewed as part of the intellectual merit and/or broader impacts of the proposal.
Proposers who feel that the plan cannot fit within the two page limit may use part of the 15-page Project Description for additional data management information.

Does not supersede specialized solicitation requirements regarding data management plans.

Data management requirements specific to the Directorate, Division, Office or other unit are available at: http://www.nsf.gov/bfa/dias/policy/dmp.jsp

FastLane will not permit submission of a proposal that is missing a data management plan.
Revised NSB-approved Merit Review Criteria

- Retain current Merit Review Criteria
- Governing principles:
  - Project of high quality and potentially transformative
  - NSF projects, in the aggregate, should contribute more achieving societal goals, i.e. Broader Impacts
    - the research itself
    - directly related activities, or
    - supported by, but are complementary to, the projects
  - Appropriate and meaningful assessment and evaluation Task force established to clarify and enhance the function of the criteria
- Task force established to clarify and enhance the function of the criteria
Societies and NGOs

Policy

- “develop a plan for citing data referenced in publications and preserving data links for the long term.”
  AMS Ad Hoc Committee on Data Stewardship Prospectus 2009
- “The following policy has been adopted for AGU publications in order to ensure that they can effectively and efficiently perform an expanded role in making the underlying data for articles available to researchers now and in the future.”
  American Geophysical Union (AGU) 1993-1996

Guidelines

- Interagency Data Stewardship/Citations/provider guidelines
  The Federation of Earth Science Information Partners (ESIP) 2012
- Digital Curation Centre in Edinburgh
  http://www.dcc.ac.uk/resources/how-guides/cite-datasets
HOW DO WE CHANGE THE CULTURE?
“Data publication is becoming increasingly important to the scientific community, as it will provide a mechanism for those who create data to receive academic credit for their work and will allow the conclusions arising from an analysis to be more readily verifiable, thus promoting transparency in the scientific process”

*Citation and Peer Review of Data: Moving Towards Formal Data Publication*
Bryan Lawrence, Catherine Jones, Brian Matthews, Sam Pepler, Sarah Callaghan
Why Data Citation?

- Transparency, access, and validation
- Accreditation - contributors and re-users
- Demonstrate impact of data as well as contributors - citation metrics
- Acceptance as a citable research reference equivalent to peer reviewed publication (new NSF guidelines)
- Critiques of the data or inspiration for use
- Encourages preservation
- Encourages the development of data discovery services
- Broadening participation
Last year at this time:
A Challenge to the Workshop:

Give us at least three ways to answer this hypothetical question a year from now:

“You told us that this new cyberinfrastructure investment would transform both the practice of science and engineering itself and lead to significant advances in knowledge and understanding – advances that would not have happened otherwise – can you give us some concrete examples of this?”
Question to Bridging Data Workshop

- What are the barriers, challenges and standards?
- How can they be addressed effectively?
- What can NSF do to affect the needed change?
- What is the role of journals, societies, and other kinds of organizations?
- How can the U.S. play a leadership role?
- How can case studies of the benefits of data citation be best documented?
Where discoveries begin

Thanks