Data Infrastructure Building BlockS (DIBBS) NSF 12-557

Dane Skow
Program Director
Data Cluster
Office of Cyberinfrastructure
Office of the Director

National Science Foundation

The DIBBS Team

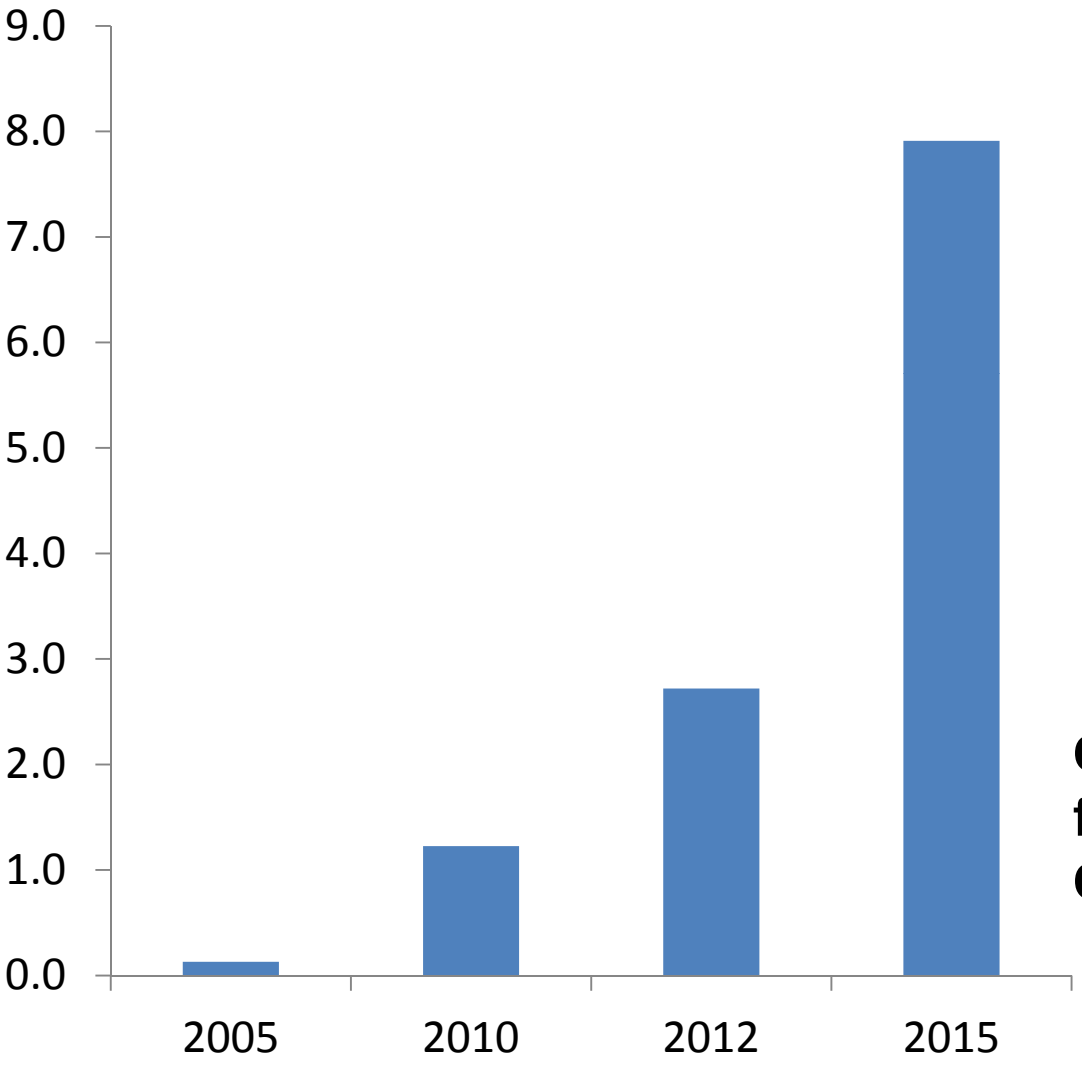
- Dane Skow, NSF – OCI
- Robert Chadduck, NSF – OCI
- Mimi McClure, NSF - OCI
- Bonnie Mountain, NSF – OCI
- Thyagarajan Nandagopal , NSF – CISE

- (others yet to be confirmed)

Outline

- Welcome – Alan Blatecky
- Webinar – Dane Skow
  - Background
    - Data Deluge
    - Research Opportunities and Challenges
    - DIBBS Program in Context
  - DIBBS Solicitation
    - Scope
    - Research Thrusts
    - Proposal Types and Deadlines
    - Proposal preparation and submission
    - Proposal Review Process
  - Q & A – Please email your questions to dibbs@nsf.gov

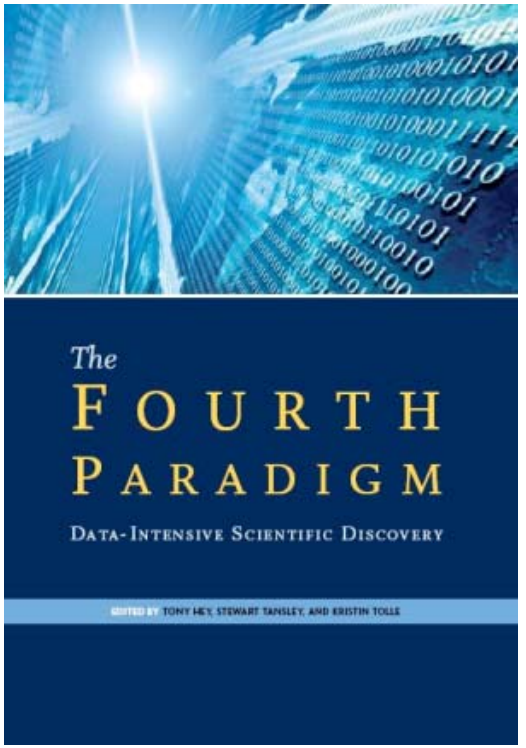
Data Deluge



Current rate of growth fills the Library of Congress every 10 sec.

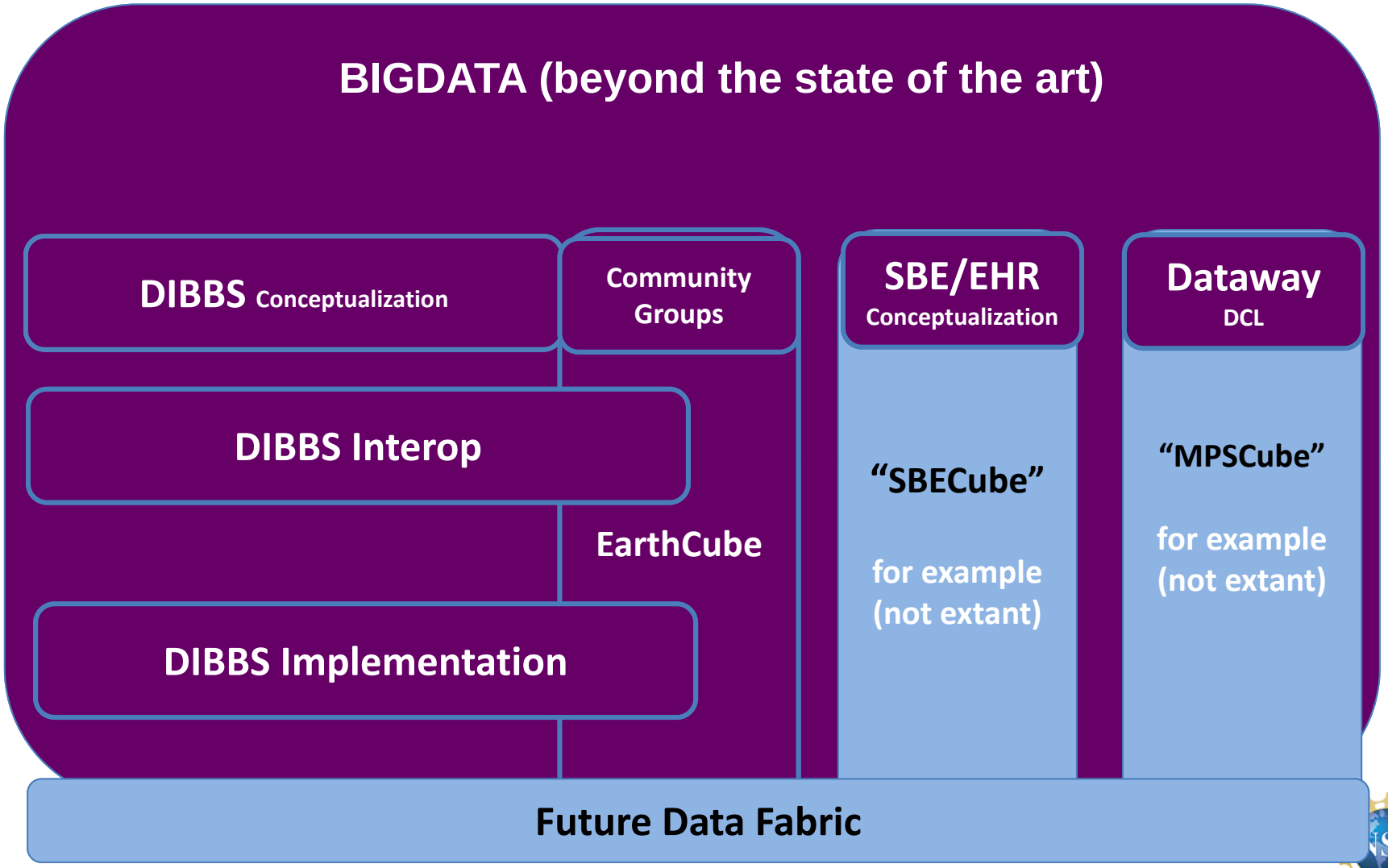
Source: EMC/IDC Digital Universe Study, 2011

Dealing with Data



http://www.sciencemag.org/site/special/data/ http://www.economist.com/node/15579717

NSF Data Solicitations



Data Web Forum (DWF)

- A Concept Paper proposing the creation of a Data Web Forum for working on data exchange issues has just been published at http://www.cni.org/wp-content/uploads/2012/06/DataWebForum_Concept_Paper.pdf

- This paper is likely of interest to participants in the webinar and you are encouraged to read it and comment.

DIBBS Solicitation

- DIBBS seeks proposals that bring into practice leading-edge tools and techniques and build community consensus with three focus areas:
  - Conceptualization Awards
    - 8-15 one year awards (up to $1.5M pool)
  - Interop Awards
    - Up to 5 up to 3 year awards (up to $1.5M each total)
  - Implementation Awards
    - Up to 4 up to 5 year awards (up to $8M each total)

Conceptualization Awards

Potential Conceptualization proposals include, but are not limited to:

- Community building and requirements gathering
- Development of teams of domain and technology experts prepared to address data infrastructure challenges
- Proof of concept trials with user communities using state of the art research results to produce/improve production data services

(8-15 awards expected, up to $1.5M pool)

Interop Awards

Potential Interop proposals include, but are not limited to:

- Merger of multiple disparate data resources into a coherent system.
- Testbeds for production testing of interoperation of state of the art solutions with current infrastructure
- Development of interoperation standards and techniques which dramatically improve the exchange, discovery, cataloging and/or curation of data
- International efforts to share data resources across geographic, network, or political divides
- Implementation testbeds of interoperation of medium scale data infrastructure building blocks

(5 awards expected, up to $1.5M each total, up to 3 yrs)

Implementation Awards

Potential implementation proposals include, but are not limited to:

- Implementation of leading edge technologies to build data infrastructure building blocks
- Development of necessary components for the construction of a national scale data infrastructure, for example, but not limited to
  - Global system of data identifiers
  - Data interchange tools and standards
  - Data management and access technologies

Preferably, but not limited to, interoperable with existing DataNet Partners (4 awards expected, up to $8M each total)

What proposals are good fits for the DIBBS solicitation?

- The focus of this solicitation is on creation of data infrastructure building blocks which can be replicated and combined to build a rich, interoperable, (inter)national data infrastructure
- Solutions which can be composed to scale from the laptop to the national resource level.
- Collaborative proposals between universities and commercial partners
- Community building exercises to help gather requirements and mobilize the necessary set of domain partners to construct appropriate solutions
- Testbeds and proof-of-concept demonstrations of state of the art technology

What proposals are not good fits for the DIBBS Solicitation?

- Proposals that focus primarily on
  - Developing basic computer science or domain specific advances in handling big data (they belong in BIGDATA)
  - Developing highly customized resources that are not readily reuseable
  - Isolated facilities intended for standalone use only
  - Monolithic solutions which are not modular and don't facilitate composition with other services

Proposal Submission and Review

- All proposals are reviewed by panels according to
  - Standard NSF merit review criteria and
  - Solicitation-specific review criteria
- In all three tracks proposals will be evaluated in part on how effective the proposed plan will be in:
  - meeting well-defined, critical data needs
  - creating new opportunities and capabilities for discovery, innovation and learning
  - improving the way science and engineering research and education are conducted
  - addressing the need for long term economic and technological sustainability beyond the end of NSF funding.

Review Criteria

Proposals in the implementation and interoperability tracks will also be assessed on how effective the plan will be in:

- addressing multiple stages in the full data management life cycle;
- developing new tools and capabilities for learning that integrate research and education;
- providing for community input and participation in all phases
- ensuring vigorous and comprehensive evaluation and assessment of all aspects of the project.
- providing the required cyberinfrastructure resources and capabilities;
- enhancing the value of existing cyberinfrastructure capabilities or services to the community;
- providing an appropriate range of expertise in cyberinfrastructure, library and archival sciences, computer and information sciences, and domain sciences;
- serving a diverse user base;
- ensuring active participation by a diverse range of individuals (including women and members of underrepresented groups), organizations, and sectors;
- serving as an effective partner in an interoperable network of organizations; and
- providing a management plan for effective leadership with clear lines of authority, responsibility, accountability, community and user responsiveness, and the ability to adapt to new opportunities and technologies.

Evaluation Plan

Each interop and implementation proposal should include a plan to evaluate the techniques and technologies developed, using, for example,

- Applications of the technology to specific domains
- Assessment of efficiency (e.g. cost, performance, scale, …)
- Operations metrics (e.g. reliability, operating cost/energy,…)
- Customer satisfaction

The evaluation plan should be appropriate for:

- The nature of proposal
- The size and scope of the project

Data Management Plan

- The types of data, 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
- The types of data and data ownership supported by services produced in the course of the project.
- Policies for access and sharing
- Policies and provisions for re-use, re-distribution, and the production of derivatives
- Plans for archiving and for preservation of access

Software Sharing Plan

- The software should be freely available to researchers and educators in the non-profit sector
- The terms of availability should permit
  - The dissemination and commercialization of enhanced or customized versions of the software, or its incorporation into other software packages
  - Further development by other groups
  - Modification to the source code and sharing of modifications
- An applicant
  - Is responsible for creating the original and subsequent official versions of software
  - May consider proposing a plan to manage and disseminate the improvements or customizations of their tools and resources by others

Proposal Types and Deadlines

NSF: http://www.nsf.gov/pubs/2012/nsf12557/nsf12557.htm

- Conceptualization:
  - Due July 26, 2012 (5 p.m. proposer's local time):
  - Budgets of approximately $100,000 for 1 year
- Interop:
  - Due August 30, 2012 (5 p.m. proposer's local time)
  - Budgets up to $1.5M (total) over up to 3 years
- Implementation:
  - Due August 30, 2012 (5 p.m. proposer's local time)
  - Budgets of up to $8M (total) over up to 5 years

Current expectation is that, subject to the availability of funds, there may be another call for proposals in FY13.

How many awards are anticipated?

- Up to $41.5 million will be invested in proposals submitted in response to this solicitation, subject to availability of funds, during FY 2012 and FY 2013.
- An estimated fifteen to twenty projects will be funded by NSF during FY 2012 and FY 2013, subject to availability of funds.

How does one apply?

- Follow the instructions provided in:
  - The solicitation http://www.nsf.gov/pubs/2012/nsf12557/nsf12557.htm
  - The Proposal and Award Policies and Procedures Guide http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/
- Consult
  - FastLane FAQ and Grants.gov FAQ: www.fastlane.nsf.gov
  - Your institution's Sponsored Research Office

Questions and Answers

- We will answer selected questions sent through email
- Answers to all common questions will be included in the FAQ (to be posted soon)
- Further Questions?
  - Email dibbs@nsf.gov

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 etc., please contact: dibbs@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 later 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