Data and Informatics Implementation

Advisory Committee to the Director Meeting

December 7, 2012

Lawrence A. Tabak, DDS, PhD

Deputy Director, NIH

Department of Health and Human Services

Charge to the Working Group

- The Advisory Committee to the Director (ACD) Data and Informatics Working Group (DIWG) will provide the ACD and the NIH Director with expert advice on the management, integration, and analysis of large biomedical datasets. The DIWG will address the following areas:

  - Research data spanning basic science through clinical and population research

  - Administrative data related to grant applications, reviews, and management

  - Management of IT at the NIH

Membership

Name	Institution	Title
David DeMets, Co-Chair	University Wisconsin - Madison	Professor, Department of Biostatistics & Medical Informatics
Lawrence Tabak, Co-Chair	NIH	Principal Deputy Director
Russ Altman	Stanford University	Professor and Chair, Department of Bioengineering
David Botstein	Princeton University	Director, Lewis-Sigler Institute
Andrea Califano	Columbia University	Chief of Biomedical Informatics
David Ginsburg, ACD Member; Chair, NCBI Needs-Assessment Panel	University of Michigan	Professor, Department of Internal Medicine; HHMI
Patricia Hurn	The University of Texas System	Associate Vice Chancellor for Health Science Research
Dan Masys	University of Washington	Affiliate Professor, Department of Biomedical Informatics and Medical Education
Jill Mesirov, Ad Hoc Member, NCBI Needs-Assessment Panel	Broad Institute	Associate Director and Chief Informatics Officer
Shawn Murphy	Harvard University	Associate Director, Laboratory of Computer Science and Assistant Professor, Department of Neurology
Lucila Ohno-Machado	University of California, San Diego	Associate Dean for Informatics, Professor of Medicine, and Chief, Division of Biomedical Informatics

Ad-hoc Membership

Name	Institution	Title
David Avrin	University of California at San Francisco	Professor and Vice Chairman, Department of Radiology
Paul Chang	University of Chicago	Professor and Vice Chairman, Department of Radiology
Christopher Chute	Mayo Clinic College of Medicine	Professor, Department of Health Sciences Research
Ted Hanss	University of Michigan Medical School	Chief Information Officer
Paul Harris	Vanderbilt University	Director, Office of Research Informatics
Marc Overcash	Emory University School of Medicine	Deputy Chief Information Officer
James Thrall	Massachusetts General Hospital, Harvard Medical School	Radiologist-in-Chief and Professor of Radiology
A. Jerome York	The University of Texas Health Science Center at San Antonio	Vice President and Chief Information Officer

Data and Informatics Working Group (DIWG): Recommendations

- Promote data sharing through central and federated catalogs

- Support development, implementation, evaluation, maintenance, and dissemination of informatics methods and applications

- Build capacity by training workforce in relevant quantitative sciences

- Develop NIH-wide data strategic plan

- Provide serious funding commitment to support these recommendations

NIH Consideration of the Recommendations

- A Pre-Implementation Team developed draft strategies for implementation for every recommendation

- IC Directors engaged in vigorous discussion of the implementation strategies at the NIH Leadership Forum

- Based on the Leadership Forum feedback, an Implementation Team refined the implementation strategies that were presented to NIH Leadership last week

Setting the Stage

A final key strategic challenge is to ensure that [the] NIH culture changes [are] commensurate with recognition of the key role of informatics and computation for every IC's mission. Informatics and computation should not be championed by just a few ICs, based on the personal vision of particular leaders. Instead, NIH leadership must accept a distributed commitment to the use of advanced computation and informatics toward supporting the research portfolio of every IC.

***Data and Informatics Working Group, ACD (June 2012 Report, p. 25)***

The Challenges We Must Solve

- We must create an adaptive and highly collaborative environment, both within NIH, and the extramural community, to enable optimal use of Big Data

- We must create a governance structure that aligns scientific leadership with resource management and oversight

- By analogy to Peer Review and support for the CSR, we must commit to a shared governance and resource plan to ensure the use and ownership of Big Data among all NIH ICs

Overarching Strategy and Goals

- Two initiatives being proposed to overcome roadblocks

    - *Big Data to Knowledge (BD2K)* – enable the biomedical research enterprise to maximize the value of biomedical data

    - *InfrastructurePlus* – create an adaptive environment at NIH to sustain world class biomedical research

- Both led by Trans-NIH Advisory Data Councils

    - Councils chaired by the NIH CIO and Chief Data Scientist (to be recruited)

    - Councils report to the NIH Director through NIH Steering Committee

Big Data to Knowledge – BD2K
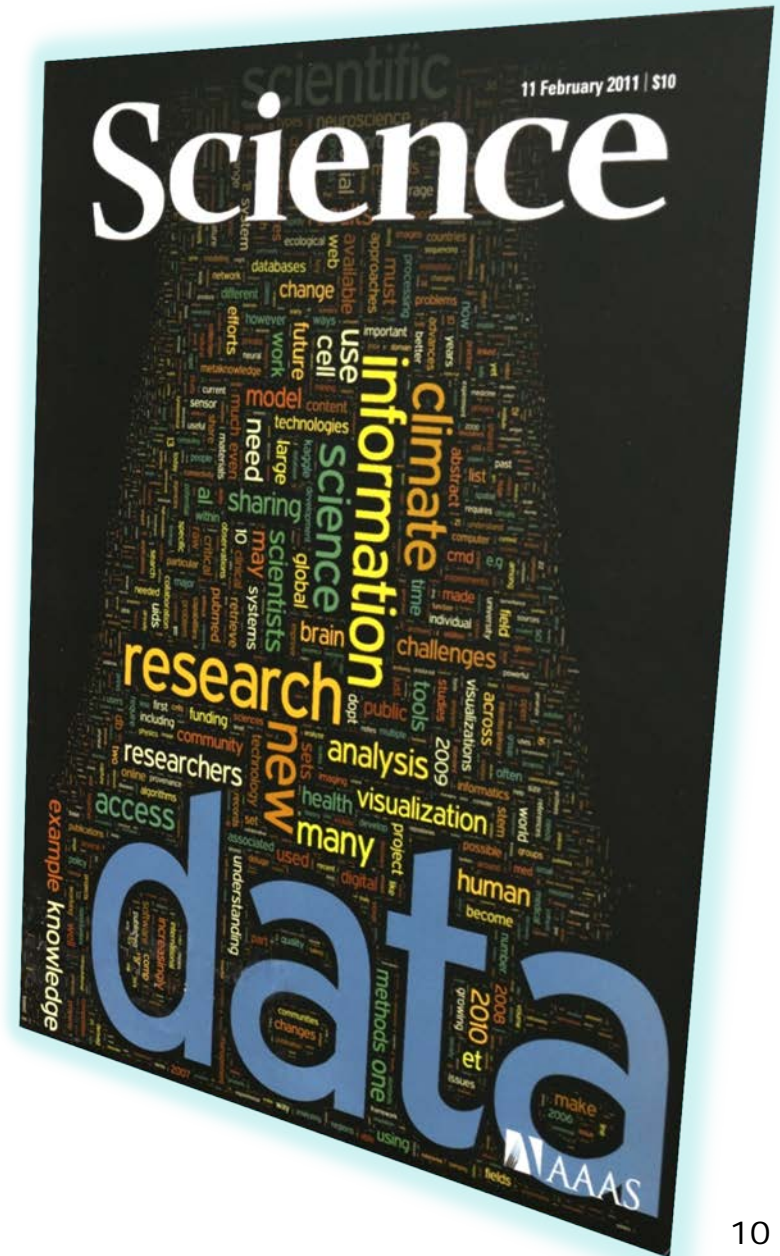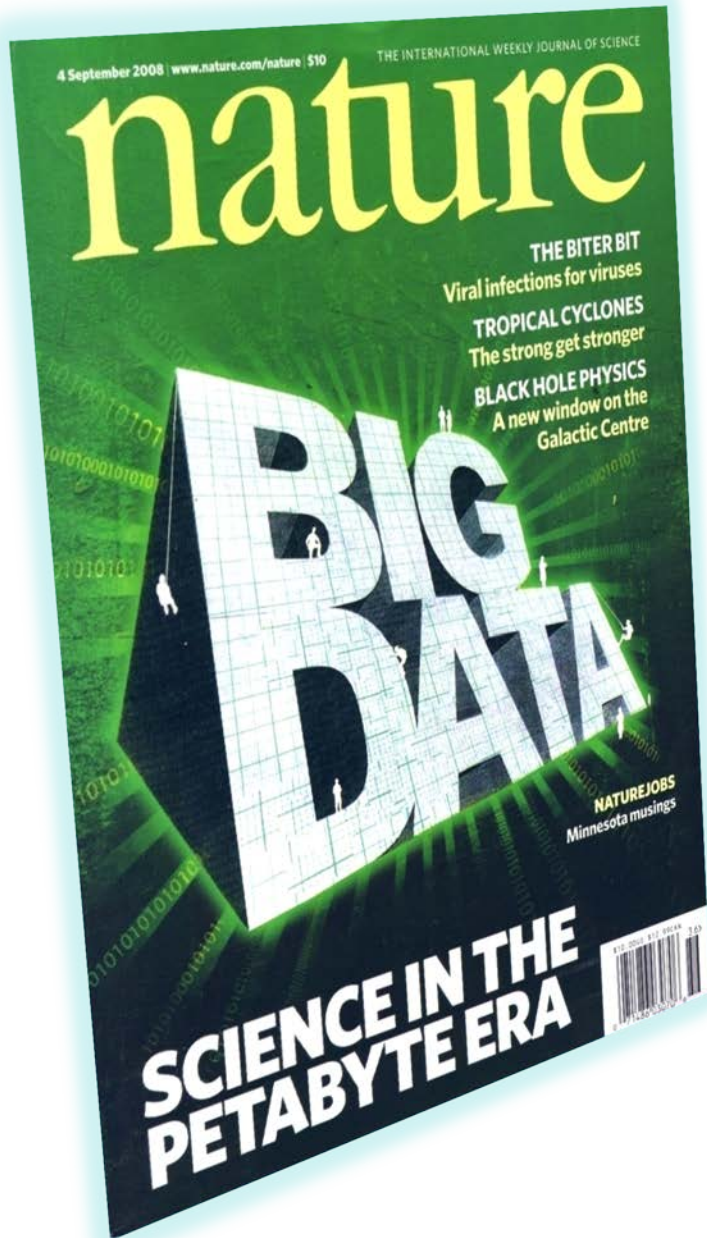
BD2K

I. Facilitating Broad Use of Biomedical Big Data

- New Policies to Encourage Data & Software Sharing

- Catalog of Research Datasets to Facilitate Data Location & Citation

- Community-based Development of Data & Metadata Standards

  - For all three areas:

  - FY2013   Workshops for planning; refine implementation plans

  - FY2014+ Research, new policies, and implementation

BD2K

II. Developing and Disseminating Analysis Methods and Software

- Software to Meet Needs of the Biomedical Research Community
  - FY2013 Workshop(s) to define software needs and update existing PARs
  - FY2014 Support for analytical software for underserved areas and data management/processing software

- Facilitating Data Analysis: Access to Large-scale Computing
  - FY2013 Investigate storage and analysis options, evaluate ongoing cloud pilots, and develop NIH policies
  - FY2014 Scientific Data Council to recommend follow-up programs

- Dynamic Community Engagement of Users and Developers
  - FY2013 Investigate innovative uses of social media

BD2K

III. Enhancing Training for Biomedical Big Data

1. Increase Number of Computationally Skilled Biomedical Trainees
   - FY2013 Workshop on computational training needs and announce availability of new training grants
   - FY2014 Implementation (supplements or new awards)

2. Strengthen the Quantitative Skills of All Biomedical Researchers
   - FY2013 Announce plan for required training in all NIH training grants
   - FY2014 Support courses and the development of new approaches (e.g., curriculum development and creation of innovative delivery approaches)

3. Enhance NIH Review and Program Oversight
   - FY2013 Convene a working group of SROs and POs to identify strategies and develop implementation plan

BD2K

IV. Establishing Centers of Excellence for Biomedical Big Data

1. Investigator-initiated Centers
   - FY2013 Issue RFA
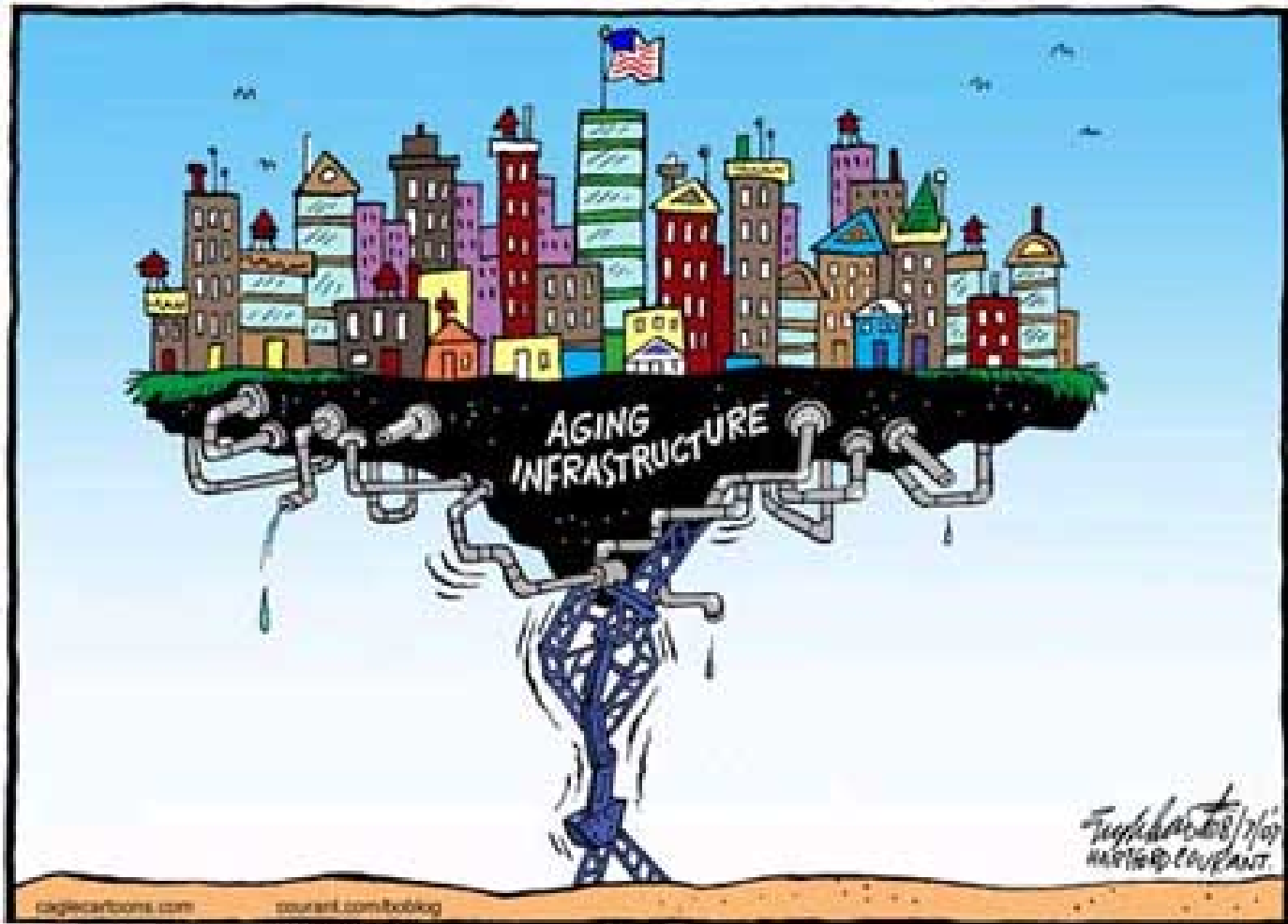   - FY2014,15 – award up to 15 centers
2. NIH-specified Centers
   - FY2013 Workshops to identify needs and issue RFA(s)
   - FY2014 award up to 2-5 centers

   Candidate Examples:
   Research data catalog and citation mechanisms, Capitalizing on EHRs, Privacy and imaging data, Genome sequence data aggregation, HumanBase
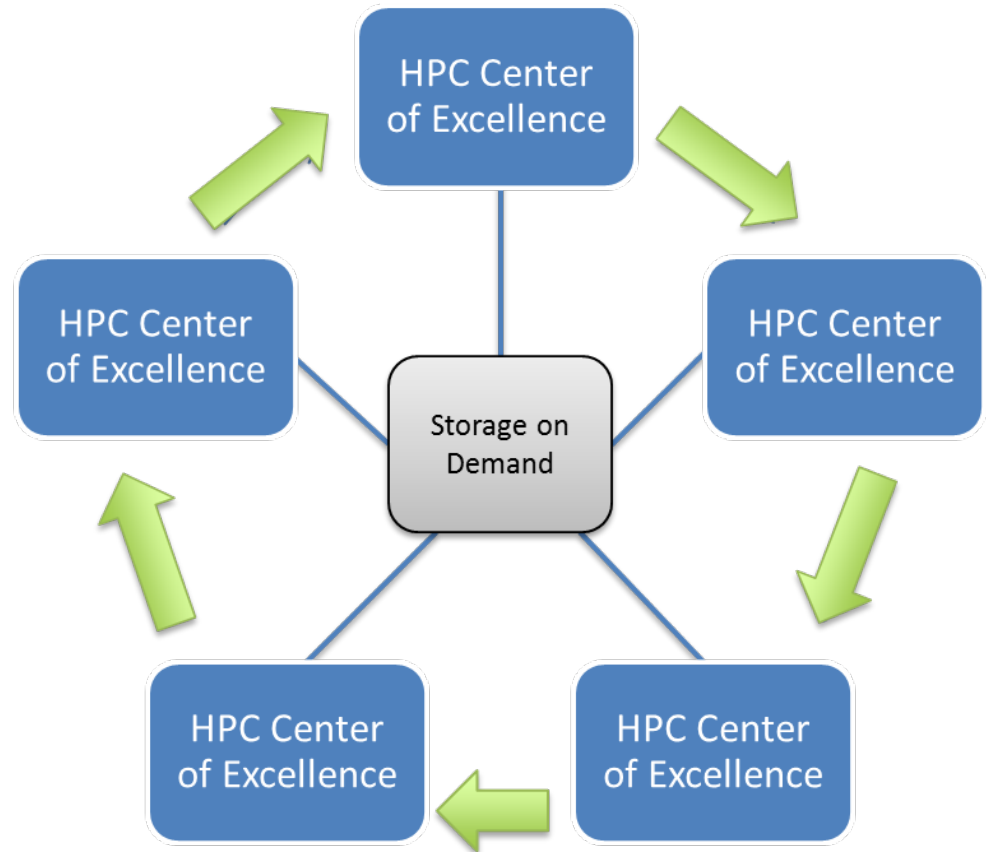
InfrastructurePlus

http://dc.streetsblog.org/wp-content/uploads/2012/09/aging_infrastructure.jpg

InfrastructurePlus

I. NIH High Performance Computational Environment

- FY2013 Activities

    - Assess NIH's current High Performance Computing capabilities and make recommendations for how to best implement a shared computational environment

- FY2014-FY2016 Activities

    - Recommend detailed implementation planning solution(s)

    - Perform migration to shared service solution(s)

InfrastructurePlus

II. Adopt Agile and Cost-Effective Hosting and Storage Approaches

- FY2013 Activities

  - Assess the NIH Data Center in order to make it more responsive and cost-competitive for IC needs

  - Assess current and projected IC computational and storage needs

- FY2014-FY2016

  - Make decision regarding transition strategy

  - Establish flexible, cost-effective, cloud contract vehicles for on-demand services

  - Adopt cloud for low risk needs

InfrastructurePlus

III. Modernize the NIH Network

- FY2013 Activities
  - Increase capacity in Internet, NIH Data Center, key locations
  - Establish new NIH Network Steering Group to guide collaborative approaches network planning

- FY2014-FY2016 Activities
  - Create new research network segment dedicated to meet the data and processing needs of the NIH's intramural program
  - Establish 100% Campus-wide wireless service and 100% Campus-wide wired & wireless network service for guests & visitors
  - Develop modern tools for a mobile and collaborative workforce (e.g., Campus-wide integrated voice, data, and video services)
  - Develop robust technologies to proactively plan/manage network capacity & performance

InfrastructurePlus

IV. Implement an Information-Rich Environment of Systems, Applications, and Tools
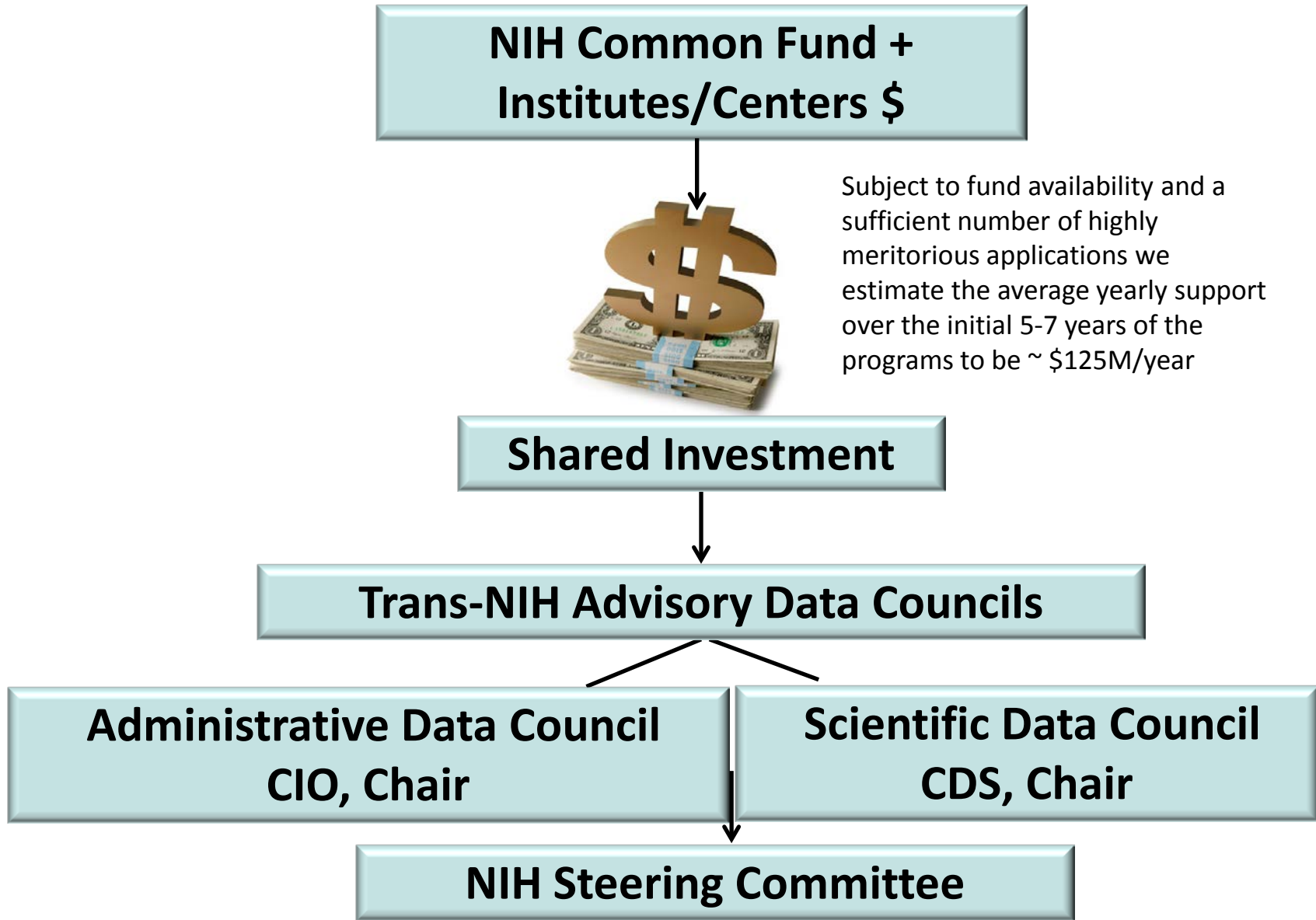
- Implement critical capabilities to support administrative and management and extramural staff, including:
  - Improved data analysis and reporting tools
  - Improved electronic grants management capabilities
  - New electronic PI BioSketch (SciENcv)
  - New contracts management application from proposal receipt to close out
  - New travel management system
  - New budget formulation and execution system
- Make critical technology upgrades and improvements for eRA and financial systems

InfrastructurePlus

V. *For Further Discussion:* **Proposed Expansion of Informatics Research in the Clinical Center**

- FY2014-FY2016

    - Expand the Laboratory for Informatics Development to create a program of pure and applied clinical informatics research

    - Initiate national search for a Director to lead the program

        - Recruit tenure-track investigators and support staff

    - Leverage BTRIS as one area of study

        - Continue as a service responsibility

        - Continue BTRIS R&D, yielding CTSA-relevant research

Scientific Leadership, Resource Management, and Governance Model

NIH Common Fund + Institutes/Centers $

Subject to fund availability and a sufficient number of highly meritorious applications we estimate the average yearly support over the initial 5-7 years of the programs to be ~ $125M/year

Shared Investment

Trans-NIH Advisory Data Councils

Administrative Data Council CIO, Chair

Scientific Data Council CDS, Chair

NIH Steering Committee

Next Steps

- Constitute Governing Boards (Advisory Data Councils) for the proposed initiatives

- Finalize plans for FY13 activities (e.g., workshops, RFAs, refine implementation plans)

- Initiate implementation plans