Advisory Committee to the Director Data and Informatics Working Group

Report to the

Advisory Committee to the Director

June 15, 2012

Drs. David DeMets and

Lawrence Tabak, Co-Chairs

"Big Data"

- Era of massive data sets

- Recent explosion of biomedical data

  - Genome sequence data

  - Public health databases

- Need for new and better ways to make the most of this data

  - Speed discovery and innovation

  - Ultimately lead to improvements in the nation's health and economy

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

Summary of DIWG Activities

- Two in-person meetings

- Seven teleconferences of the full working group

- Subgroup discussions and expert interviews

  - Research data types

  - NIH data and IT considerations

- Request for information (RFI)

  - 50 comments received

  - Redacted comments and analysis of the comments are available online: http://acd.od.nih.gov/DIWG.htm

DIWG Vision Statement

The colossal changes in technologies and methods for doing biomedical research have shifted the bottleneck in science productivity from data production to data management, communication, and data interpretation.

Given the current and emerging needs of the biomedical research community, NIH has a number of key opportunities to encourage and better support a research ecosystem that leverages data, tools, and the biomedical workforce.

Overview of Recommendations

Promote data sharing through central and federated catalogues

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

Provide a serious funding commitment to support Recommendations

Build capacity by training the workforce in the relevant quantitative sciences

Develop an NIH-wide IT strategic plan

Recommendation 1

Promote data sharing through central and federated catalogues

- 1a. Establish a minimal set of relevant data descriptions for data sharing:

  - Learn from the Google model - minimal format restrictions for data plus applications development to create new knowledge

  - Facilitate the non-expert user to easily find, access, and use data

  - Convene experts to define the metadata framework

Recommendation 1 (cont.)

Promote data sharing through central and federated catalogues

- 1b. Create catalogues and tools to facilitate data sharing:

  - Establish a centralized catalogue of data appendices

  - Link to the published literature

  - Include associated metadata as defined by the framework

Recommendation 1 (cont.)

Promote data sharing through central and federated catalogues

- 1c. Enhance and incentivize a data sharing policy for NIH funded data:

    - Update the current data sharing policy to require additional availability of data

    - Make the number of accesses/downloads from the centralized catalogue available

    - Create and provide model data use agreements to facilitate appropriate sharing of data

Recommendation 2

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

- 2a. Fund all four phases of scientific software development via appropriate, targeted mechanisms:
  - Prototyping
  - Engineering and hardening
  - Dissemination
  - Maintenance and support

Recommendation 2 (cont.)

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

- 2b. Assess how to leverage the lessons learned from the National Centers for Biomedical Computing (NCBCs):

  - The NCBCs have been a valuable engine of collaboration

  - Consider the natural evolution of the NCBCs into a more refocused activity

Recommendation 3

Build capacity by training the workforce in the relevant quantitative sciences (e.g., bioinformatics, biomathematics, biostatistics, and clinical informatics)

- 3a. Increase funding for quantitative training and fellowship awards:
  - Training of experts should grow to meet the increasing demand in this field
  - Perform a supply versus demand gap analysis
  - Develop a strategy to meet the demand

Recommendation 3 (cont.)

Build capacity by training the workforce in the relevant quantitative sciences (e.g., bioinformatics, biomathematics, biostatistics, and clinical informatics)

- 3b. Enhance review of quantitative training applications:

  - Specialized quantitative training grants are often not reviewed by those with the most relevant experience

  - Consider formation of a new study section focused on the review of quantitative science training grants

Recommendation 3 (cont.)

Build capacity by training the workforce in the relevant quantitative sciences (e.g., bioinformatics, biomathematics, biostatistics, and clinical informatics)

- 3c. Create a required quantitative component for all NIH training and fellowship awards:

  - Enable the clinical and biological scientist workforce with basic proficiency in the understanding and use of quantitative tools

  - Draw on experience of the Clinical and Translational Science Awards (CTSAs) centers in developing the curriculum for a core competency

Recommendation 4

Develop an NIH-wide IT strategic plan to be cost effective by avoiding redundancies, filling gaps, and disseminating successes to the wider NIH community

- 4a. NIH administrative data:

  - Update and share the inventory of existing analytic and reporting tools

  - Enhance coordination and sharing of resources and tools

Recommendation 4 (cont.)

Develop an NIH-wide IT strategic plan to be cost effective by avoiding redundancies, filling gaps, and disseminating successes to the wider NIH community

- 4b. NIH Clinical Center:

  - Enhance coordination of common services

  - Create a new informatics lab

  - Strengthen relationships with the CTSAs centers and the National Center for Advancing Translational Sciences (NCATS)

Recommendation 4 (cont.)

Develop an NIH-wide IT strategic plan to be cost effective by avoiding redundancies, filling gaps, and disseminating successes to the wider NIH community

- 4c. NIH IT and informatics environment:
  - Assess the current state of IT services/capabilities
  - Develop a plan for trans-NIH IT design and implementation model for IT initiatives
  - Continue to refine and expand IT governance
  - Recruit a Chief Science Information Officer (CSIO)
  - Establish an external advisory group for the NIH Chief Information Officer (CIO) and CSIO

Recommendation 5

Provide a serious, substantial, and sustained funding commitment to Recommendations 1-4

- Without a systematic and increased investment to advance computation and informatics support at the trans-NIH level and at every NIH Institute and Center, the research community will not be able to optimally use the massive amount of data that are currently being generated with NIH funding.

  - Create a sustained funding mechanism for IT

  - Motivate a culture change to recognize the key role of informatics and computation in the NIH mission

Summary of Recommendations

Promote data sharing through central and federated catalogues

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

Provide a serious funding commitment to support Recommendations

Build capacity by training the workforce in the relevant quantitative sciences

Develop an NIH-wide IT strategic plan

Questions?