# *ACD Data and Informatics Working Group*

## June 10, 2011
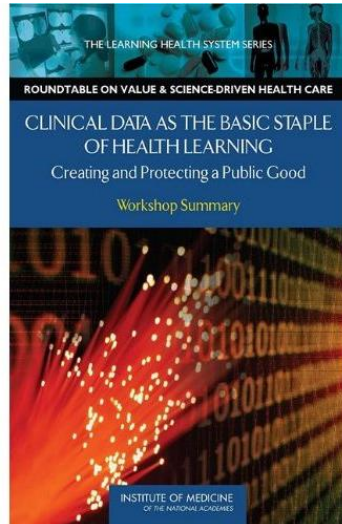
**Lawrence A. Tabak, D.D.S., Ph.D.**

**Principal Deputy Director**

**National Institutes of Health**

# Data Are A Vital Resource



**RESEARCH** ARTICLES

## The World's Technological Capacity to Store, Communicate, and Compute Information

Martin Hilbert[1]* and Priscila López[2]

We estimated the world's technological capacity to store, communicate, and compute information, tracking 60 analog and digital technologies during the period from 1986 to 2007. In 2007, humankind was able to store $2.9 \times 10^{20}$ optimally compressed bytes, communicate almost $2 \times 10^{21}$ bytes, and carry out $6.4 \times 10^{18}$ instructions per second on general-purpose computers. General-purpose computing capacity grew at an annual rate of 58%. The world's capacity for bidirectional telecommunication grew at 28% per year, closely followed by the increase in globally stored information (23%). Humankind's capacity for unidirectional information diffusion through broadcasting channels has experienced comparatively modest annual growth (6%). Telecommunication has been dominated by digital technologies since 1990 (99.9% in digital format in 2007), and the majority of our technological memory has been in digital format since the early 2000s (94% digital in 2007).

# Working Group Charge

- Provide the ACD and the NIH Director with expert advice on the management, integration, and analysis of large biomedical datasets
- The ACD WG will integrate their efforts with the
  - NCBI Needs-Assessment Panel
  - NIH IT Assessment Group

# Areas to Address

- Research data spanning basic science through clinical and population research

  - The connection and integration of large volumes of "omics" data with other large data sets including clinical and phenotypic data

  - The management and curation of these large data sets, including the use of new and emerging technologies (e.g. clouds)

  - The analysis of these integrated data sets to facilitate the development of more sophisticated predictive models of disease susceptibility and pathobiology

# Areas to Address

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

  - The management and curation of large administrative data sets, including the use of new and emerging technologies (e.g. clouds)

  - The analysis of integrated data sets to facilitate strategic planning and portfolio management

  - Where applicable, the release of this information in a 'user-friendly' manner

# Areas to Address

- Management of information technology (IT) at the NIH

  - The appropriate level of centralization of IT and/or data management

  - The identification of best practices of IT and data management for the NIH to adopt

# Proposed Membership of Working Group

| | | |
|---|---|---|
| **David DeMets, Co-Chair** | **Univ Wisconsin** | **Professor, Dept. Biostatistics & Medical Informatics** |
| Russ Altman | Stanford University | 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 | Univ of Michigan | Professor, Internal Medicine; HHMI |
| Patricia Hurn | Univ of Texas | Associate Vice Chancellor Health Science Research The University of Texas System |
| Dan Masys | Vanderbilt University | Chair of the Department of Biomedical Informatics |
| Jill Mesirov, Ad Hoc Member, NCBI Needs-Assessment Panel | Broad Institute | Associate Director and Chief Information Officer |
| Shawn Murphy | Harvard | Assistant Professor |
| Lucila Ohno-Machado | Univ of San Diego | Chief Division of Biomedical Informatics |