What's Next for Data Sharing: Insight from the NIH Experience

Jerry Sheehan
Assistant Director for Policy Development
National Library of Medicine – National Institutes of Health

SHARE In-Person Meeting
Crystal City, VA
October 14, 2014

NIH Data Sharing – Why?

- Preserve the scientific record (of which data are a growing part)
- Support portfolio management and analysis
- Speed translation of research results into knowledge, products, and procedures to improve human health.
- Assist in reproducibility/replication
- Allow further mining of collected data
- Allow aggregation of data from multiple studies
- Allow reuse of data for additional research

All aim to improve the quality and efficiency of biomedical research.

NIH Data Sharing Policy - 2003

[http://sharing.nih.gov]

. . .investigators submitting an NIH application seeking $500,000 or more in direct costs in any single year are expected to include a plan for data sharing or state why data sharing is not possible.

- Reviewers not to factor the proposed data-sharing plan into the determination of scientific merit or priority score
- Few implementation details specified
  - Data defined as “Recorded factual material commonly accepted in the scientific community as necessary to document and support research findings”
  - Timely release and sharing considered “no later than the acceptance for publication of the main findings from the final data set”
  - Applicant to determine where to share data, rules for access, etc.

Key Elements of Data Sharing Policy

http://grants.nih.gov/grants/sharing_key_elements_data_sharing_plan.pdf

- **WHO** must share data and WHO will have access to the data?
  - Other researchers?
  - General public?
- **WHAT** data will be shared? WHAT limitations on access (e.g., privacy)
  - Types of data collected (genetic, clinical, images)
  - Where along data processing pipeline
  - Data documentation (descriptors, metadata)
  - Data standards, vocabularies used
- **WHERE** will data be located, deposited, made available?
  - Existing repository, newly established one?
  - Shared directly by investigator?
- **WHEN** will data be shared? WHEN may it be accessed?
  - First available prior to publication or upon publication
  - Data maintenance after end of award
- **HOW** will researchers locate and access data? HOW may they use it?
  - Is data discoverable and what steps needed to access it (e.g., approvals?)
  - Is proposed use consistent with, e.g., informed consent

NIH Data Sharing Policies

This table lists data sharing policies in effect at NIH. It includes policies at the NIH, IC, division, and program levels that apply to broad sets of investigators and data. Individual requests for applications (RFAs) and program announcements (PA) may specify other requirements or expectations for data sharing that apply to specific projects.

Show 50 entries

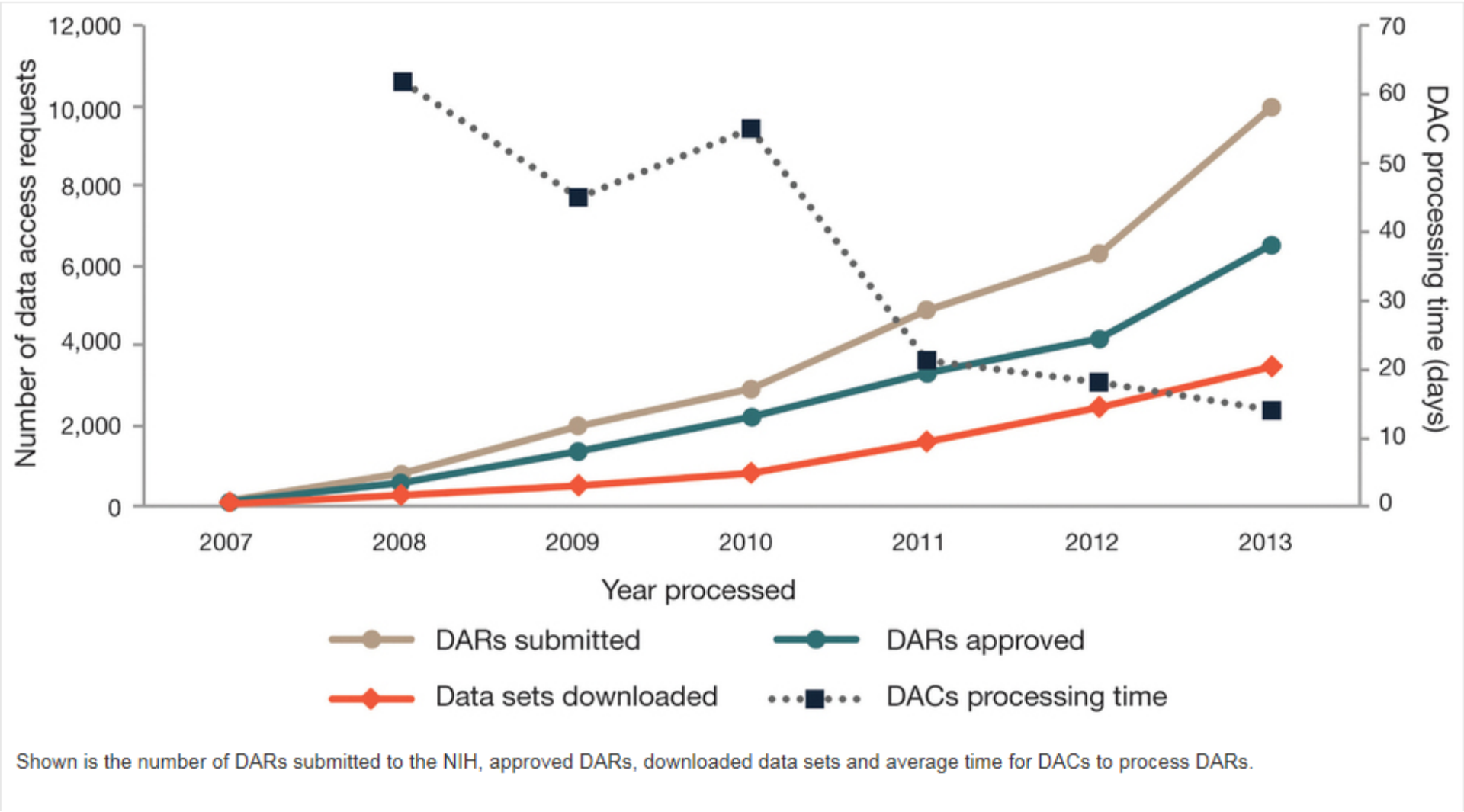
Search:

IC	Data Sharing Policy Name	Description of Data Sharing Policy	Repositories
ORDR	The Collaboration, Education, and Test Translation (CETT) Program's Guidelines for Data Collection and Sharing	Specifies that de?identified clinical data will be submitted and stored at the NIH for future distribution for research purposes. To facilitate the widest access to data, CETT Collaborative teams agree to the following principles: a) follow de-identification procedures defined within the GWAS policy b) develop procedures and educational/informational documents and c) de-identified clnical data will be submitted and stored at the NIH for future distribution for research purposes.	dbGaP
NINDS	Sharing Data via the Federal Interagency Traumatic Brain Injury Research (FITBIR) Informatics System	Investigators submitting FITBIR data are expected to: a) provide descriptive information about their studies, b) sumbit coded genotypic and phenotypic data to the FITBIR Informatics System; and c) submit a data submission for providing assurance that all data are submitted to the DOD and the NIH in accord with applicable laws and regulations, and that the identities of research participants will not be disclosed to the FITBIR Informatics System.	FITBIR
NIH	NDAR Grantees Data Sharing Policy	All data resulting from this autism-related NIH-funded research involving human subjects are expected to be submitted to the National Database for Autism Research (NDAR), along with appropriate supporting documentation to enable efficient use of the data.	NDAR
NIH	NIH Data Sharing Policy	Expects investigators seeking more than $500K in direct support in any given year to submit a data sharing plan with their application or to indicate why data sharing is not possible.	No specific repository listed
NIH	NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS)	Expects all investigators who receive NIH support to conduct genome-wide analysis of genetic variation in a study population to submit to the NIH GWAS data repository descriptive information about their studies for inclusion in an open access portion of the NIH GWAS data repository. Strongly encourages the submission of curated and coded phenotype, exposure, genotype, and pedigree data, as appropriate, to the NIH GWAS data repository as soon as quality control procedures have been completed at the local institution. These detailed data will be made available through a controlled access process according to the GWAS Data Access procedures.	dbGaP
NIH	NIH Policy on Deposit of Atomic Coordinates into Structural Databases	NIH policy requires that atomic coordinates from X-ray crystallographic and nuclear magnetic resonance experiments that were supported by NIH grants be deposited into the appropriate structural database at the time of submission of a research article drawing conclusions from these data.	Protein Data Bank

NIH Genome Wide Association Study (GWAS) Policy Details

Who	Investigators (intramural and extramural) who proposes a study that include GWAS
What	• Include a plan for sharing genotype and phenotype data GWAS data to or explain why not. • Participant-level genotype data, phenotype data, plus protocol
Where	Deposit data in designated repository – dbGAP
When	• Data to be submitted as soon as possible after cleaning • Data available to other investigators upon submission • Publication exclusivity for 12 months after deposit of data.
How	• General descriptive data are available to the public • Access to participant-level data (de-identified) for secondary research projects must be approved by a Data Access Committee (DAC) • Investigators must agree to security and confidentiality requirements • Secondary users expected to acknowledge contributing investigators and funding organizations for original studies
Why	Facilitate broad access to NIH-supported GWAS data to speed translation of basic genetic research into therapies, products, and procedures that benefit the public health

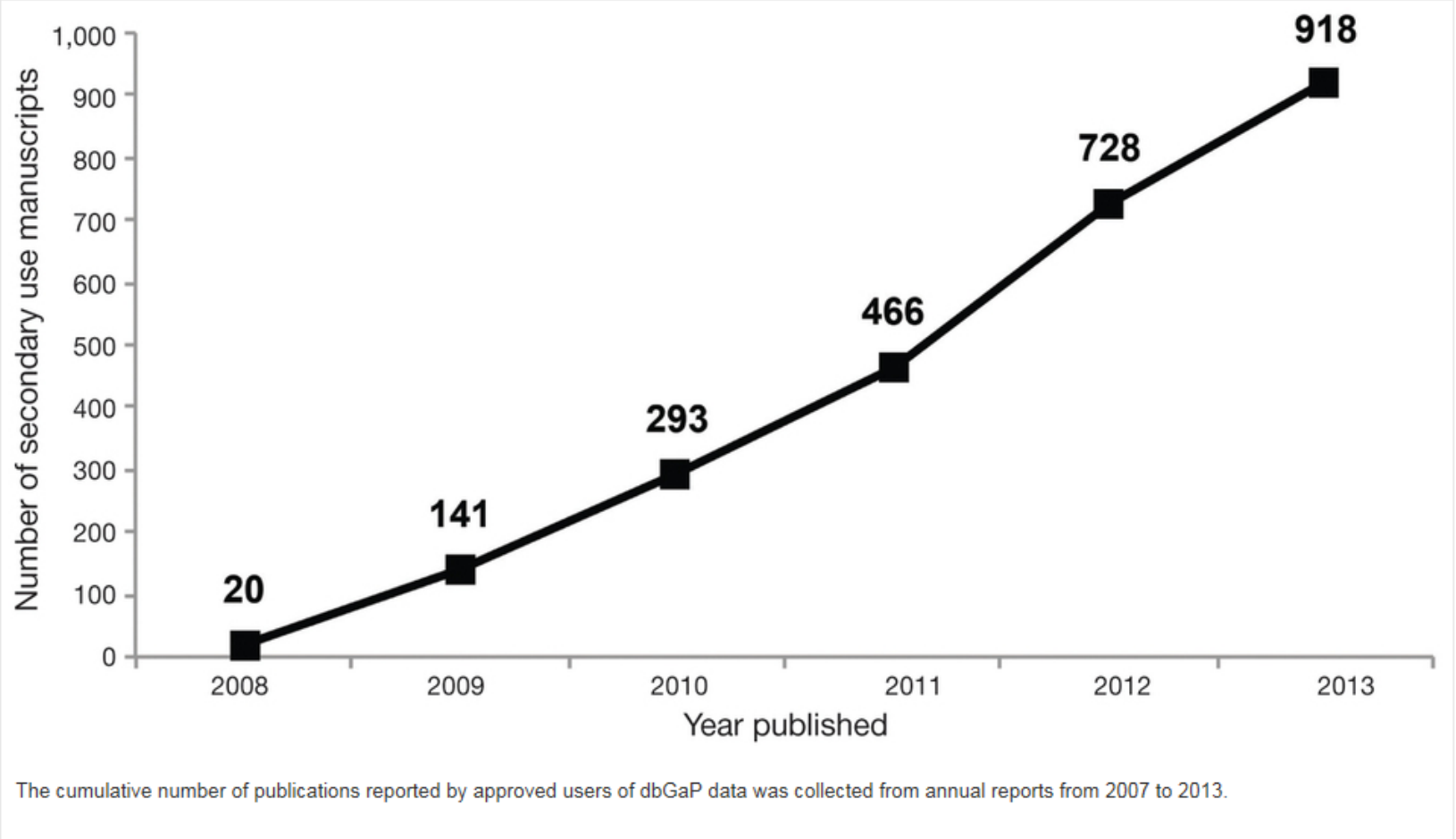
Increasing requests for GWAS Data



Shown is the number of DARs submitted to the NIH, approved DARs, downloaded data sets and average time for DACs to process DARs.

Source: Paltoo, et al., "Data use under the NIH GWAS Data Sharing Policy and future directions," Nature Genetics 46, 934-938. Published online 27 Aug 2014. doi:10.1038/ng.3062.

Publications based secondary use of GWAS data



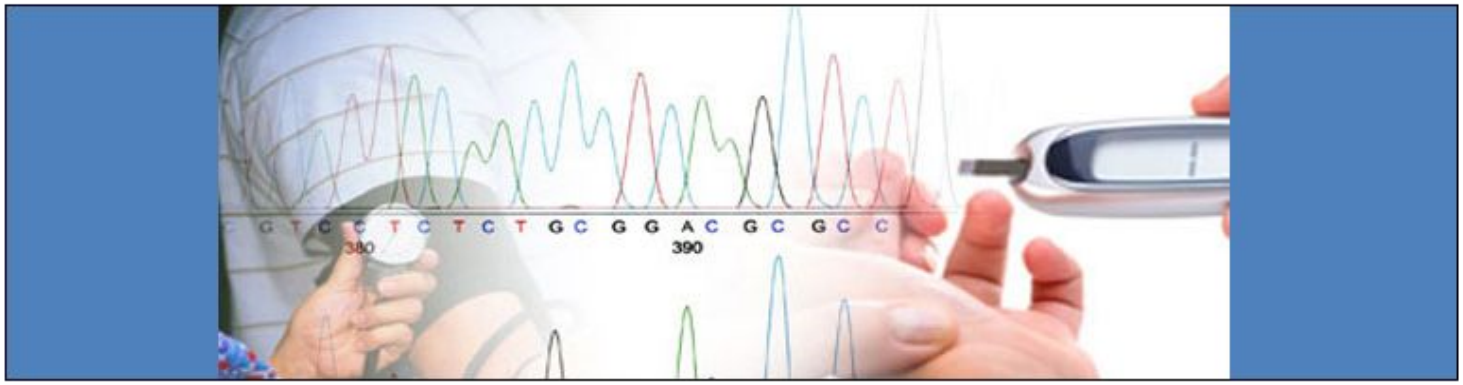
The cumulative number of publications reported by approved users of dbGaP data was collected from annual reports from 2007 to 2013.

Source: Paltoo, et al., "Data use under the NIH GWAS Data Sharing Policy and future directions," Nature Genetics 46, 934-938. Published online 27 Aug 2014. doi:10.1038/ng.3062.

NIH Genomic Data Sharing (GDS)

Announced August 27, 2014

Home
Policy
Policy Oversight
Researchers
Institutions & IRBs
Data Repositories
FAQs
Related Resources
Subscribe to the GDS LISTSERV
Facts & Figures



Introduction

Genomic research advances our understanding of factors that influence health and disease, and sharing genomic data provides opportunities to accelerate that research through the power of combining large and information-rich datasets. To promote sharing of human and non-human genomic data and to provide appropriate protections for research involving human data, the National Institutes of Health (NIH) issued the Genomic Data Sharing (GDS) Policy on August 27, 2014, in the *NIH Guide Grants and Contracts* (available at http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html), and in the *Federal Register* (available at https://federalregister.gov/a/2014-20385) on August 28, 2014. The GDS Policy is also available at http://gds.nih.gov/03policy2.html.

The GDS Policy applies to all NIH-funded research (e.g., grants, contracts, and intramural research) that generates large-scale human or non-human genomic data, regardless of the funding level, as well as the use of these data for subsequent research. Large-scale data include genome-wide association studies (GWAS), single nucleotide polymorphisms (SNP) arrays, and genome sequence, transcriptomic, epigenomic, and gene expression data. Supplemental Information to the GDS Policy (available at http://gds.nih.gov/pdf/supplemental_info_GDS_Policy.pdf) provides examples of genomic research projects that are subject to the Policy and the timeline for submission and sharing of data from such projects.

The GDS Policy takes effect for grant applications with due dates on or after January 25, 2015, for contracts submitted on or after January 25, 2015, and for intramural research projects generating genomic data on or after January 25, 2015.

Questions about the Policy can be e-mailed to GDS@mail.nih.gov.

Clinical Trial Data Policy

Food and Drug Administration Amendments Act of 2007

Who	Responsible parties (sponsors or designated PI's) for Applicable Clinical Trials of FDA-regulated drugs and devices
What	• Specified registration information for all ACTs (protocol description) • Summary results information, including adverse event information for trials of products that are approved, cleared, licensed by FDA
Where	ClinicalTrials.gov
When	• Register not later than 21 days after enrolling first participant • Results not later than 1 year after date of final data collection for the primary outcome • Information posted not later than 30 days after submission (except for trials of unapproved/uncleared devices)
How	• All information is publicly accessible and searchable (no participant level data • Data must be deposited via an organizational account
Why	• To enhance enrollment and provide a mechanism to track progress of clinical trials • To provide more complete results information and enhance access to and understanding of the results

ClinicalTrials.gov

A service of the U.S. National Institutes of Health

ClinicalTrials.gov is a registry and results database of publicly and privately supported clinical studies of human participants conducted around the world. Learn more about clinical studies and about this site, including relevant history, policies, and laws.

Find Studies | About Clinical Studies | Submit Studies | Resources | About This Site

ClinicalTrials.gov currently lists 176,132 studies with locations in all 50 states and in 187 countries.

Text Size

Search for Studies

Example: "Heart attack" AND "Los Angeles"

Search

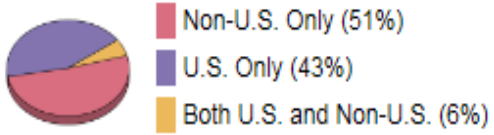
Advanced Search | See Studies by Topic

See Studies on a Map

Search Help

- How to search
- How to find results of studies
- How to read a study record

Locations of Recruiting Studies



Total N = 33,799 studies

Data as of October 07, 2014

- See more trends, charts, and maps

For Patients & Families

- How to find studies
- See studies by topic
- Learn about clinical studies
- Learn more...

For Researchers

- How to submit studies
- Download content for analysis
- About the results database
- Learn more...

For Study Record Managers

- Why register?
- How to register study records
- FDAAA 801 Requirements
- Learn more...

Learn More

- ClinicalTrials.gov Online Training
- Glossary of common site terms

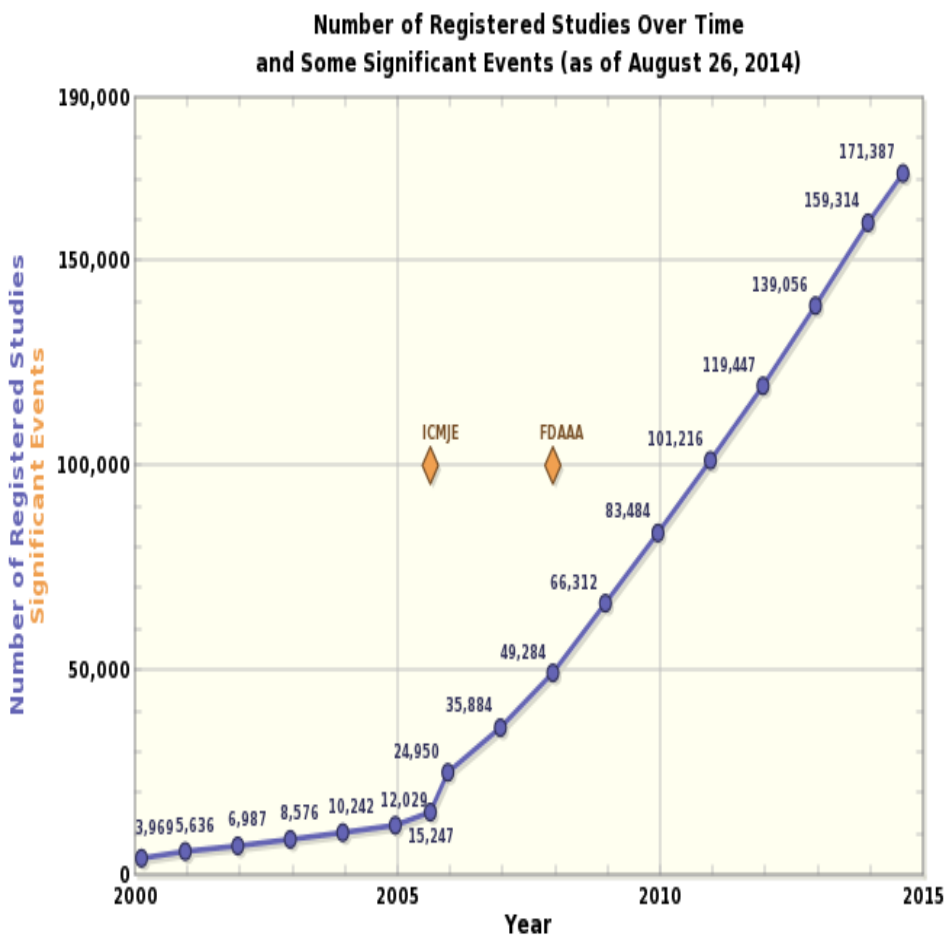
For the Press

Using our RSS Feeds

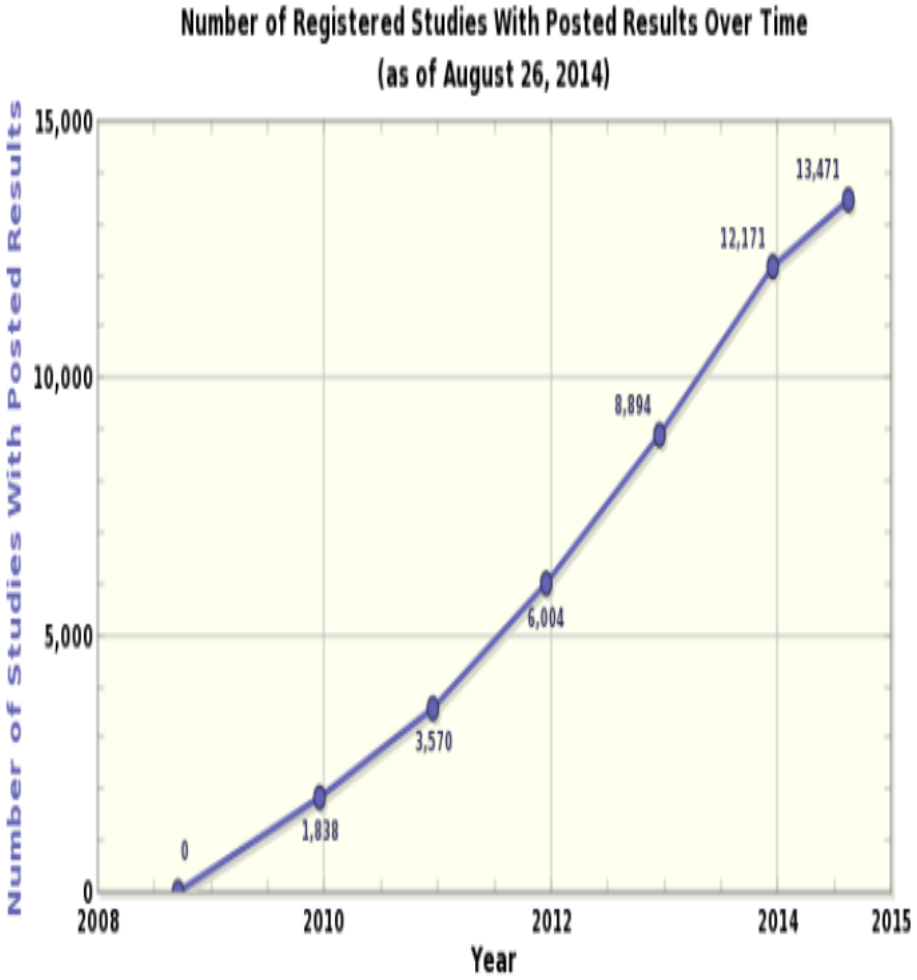
112+ million page views per month
57,000 unique visitors per day

HOME | RSS FEEDS | SITE MAP | TERMS AND CONDITIONS | DISCLAIMER | CONTACT NLM HELP DESK

Increased access to clinical trial data



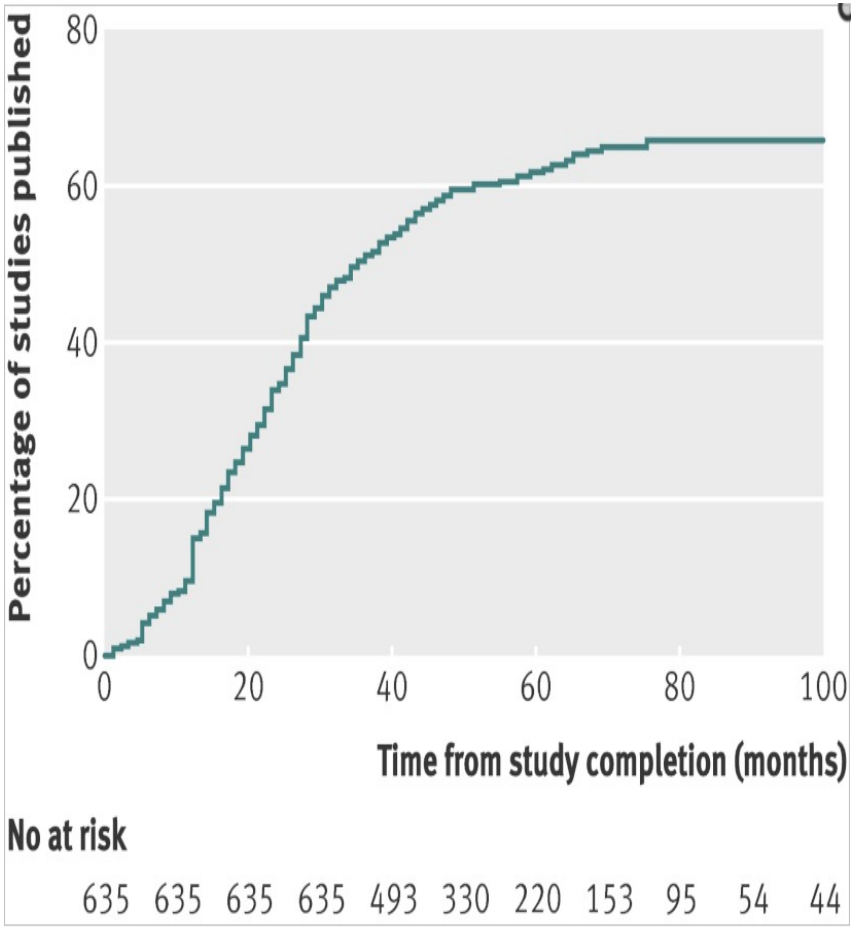
Source: http://ClinicalTrials.gov



Source: http://ClinicalTrials.gov

Insight into Clinical Research

- In 2012 alone, 1,300-3,700 trials not covered by federal human subjects protections regulations
- Of 110 trials with results posted on ClinicalTrials.gov and cited in MEDLINE
  - 16 (15%) inconsistent description of Primary Outcome Measure
  - 22 (20%) inconsistent value for Primary Outcome Measure
- Characteristics of registered trials
  - Many small trials; many single arm trials;
  - Heterogeneity in study design (blinding, DMCs) across different fields
- One-third of NIH-funded trials remained unpublished after a median of 51 months



Source: Ross, et al., Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis, BMJ 2012; 344:d7292

Clinical Trial Data – Next Steps

- Publication of Notice of Proposed Rulemaking
  - Clarify statutory requirements
  - Address key issues identified in FDAAA: which trials, what information, timelines
- IOM study: Guiding Principals for Clinical Trial Data Sharing - *forthcoming*

Enhancing NIH Data Sharing Policies: OSTP Memo (22 February 2013)

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

1. Policy Principles

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

Scientific research supported by the Federal Government catalyzes innovative breakthroughs that drive our economy. The results of that research become the grist for new insights and are assets for progress in areas such as health, energy, the environment, agriculture, and national security.

Access to digital data sets resulting from federally funded research allows companies to focus resources and efforts on understanding and exploiting discoveries. For example, open weather

OSTP Memo: Principles for Access to Scientific Data

- Maximize free access while
  - Protecting privacy and confidentiality
  - Recognizing intellectual property rights
  - Balancing costs and benefits of long-term preservation
- Ensure submission of data management plans
- Allow inclusion of costs in applications for funding
- Ensure appropriate evaluation of DMPs
- Monitor compliance by intramural and extramural investigators
- Deposit data in public repositories, where possible
- Cooperate with the private sector
- Develop approaches for citation and attribution of data sets
- Support training, education and workforce development
- Assess long-term needs for preservation and options for repositories

NIH Plan aims to meet these objectives

NIH Big Data To Knowledge (BD2K) Initiative

1. Facilitate Broad Use of Biomedical Big Data
2. Develop and Disseminate Analysis Methods and Software for Biomedical Big Data
3. Enhance Training for Biomedical Big Data
4. Establish Centers of Excellence for Biomedical Big Data

NEWS & EVENTS

News & Events

News Releases
Events
Videos
Images
Social Media & Outreach

NIH News in Health
NIH Research Matters
NIH Record

For Immediate Release: Thursday, October 9, 2014, 10:30 a.m. EDT

NIH invests almost $32 million to increase utility of biomedical research data

Wide-ranging National Institutes of Health grants announced today will develop new strategies to analyze and leverage the explosion of increasingly complex biomedical data sets, often referred to as Big Data. These NIH multi-institute awards constitute an initial investment of nearly $32 million in fiscal year 2014 by NIH's Big Data to Knowledge (BD2K) initiative, which is projected to have a total investment of nearly $656 million through 2020, pending available funds.

With the advent of transformative technologies for biomedical research, such as DNA sequencing and imaging, biomedical data generation is exceeding researchers' ability to capitalize on the data. The BD2K awards will support the development of new approaches, software, tools, and training programs to improve access to these data and the ability to make new discoveries using them. Investigators hope to explore novel analytics to mine large amounts of data, while protecting privacy, for eventual application to improving human health. Examples include an improved ability to predict who is at increased risk for breast cancer, heart attack and other diseases and condition, and better ways to treat and prevent them.

Institute/Center

NIH Office of the Director (OD)

Contact

NIH Office of Communication and Public Liaison
301-496-5787

Related Links

NIH BD2K Awardees and Project Summaries

BD2K Website

Mulitmedia

NIH Big Data to Knowledge (BD2K) — Infographic

Subscribe

Receive NIH news releases by e-mail

Interagency cooperation: CENDI

- Interagency group of senior federal STI managers
  - 12 member agencies manage > 97% of annual federal R&D budget
- Interests for FY15 include
  - Data management
  - Data repositories and infrastructure
  - Public access to publications

Many remaining challenges

- Identification of high-value data -- and approaches for doing so
- Metadata -- to support discovery and reuse
- Mechanisms for data discovery
- Repositories (centralized, distributed) -- and criteria for trusted repositories
- Elements of a data management plan
- Mechanisms for citation and attribution
- Approaches for linking data, analytical software, publications and other research objects
- . . .

All require broad consultation and cooperation

Thank you

NIH Data Sharing Policy: http://sharing.nih.gov

Genomic Data Sharing Policy: http://gds.nih.gov/

Clinical Trials Data Sharing: http://clinicaltrials.gov

NIH Big Data To Knowledge: http://bd2k.nih.gov

CENDI: http://www.cendi.gov/

NIH Data Sharing Policies:
http://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html

NIH Data Sharing Repositories:
http://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html