

The National Library of Medicine and the National Institutes of Health Partnership in Accelerating Data Driven Discovery

Strategic Plan 2017-2027

How the NLM creates Trust in Data Patricia Flatley Brennan, RN, PhD NLM Director



#### Overview

- NLM Strategic Plan
- NIH Strategic Plan for Data Science
- How NLM is preparing for a future of data-driven discovery
- NIH Training and Policy Options







#### Clinical Trials.gov

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

## 





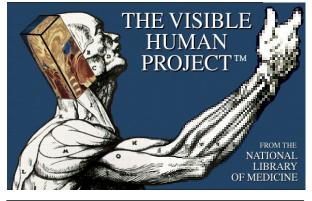




























**NCBI** 



**HISTORY** 

OF MEDICINE



TOXNET TOXICOLOGY DATA NETWORK









## NLM Strategic Plan Transforming Data into Knowledge



Accelerate discovery and advance health through datadriven research



Reach more people in more ways through enhanced dissemination and engagement



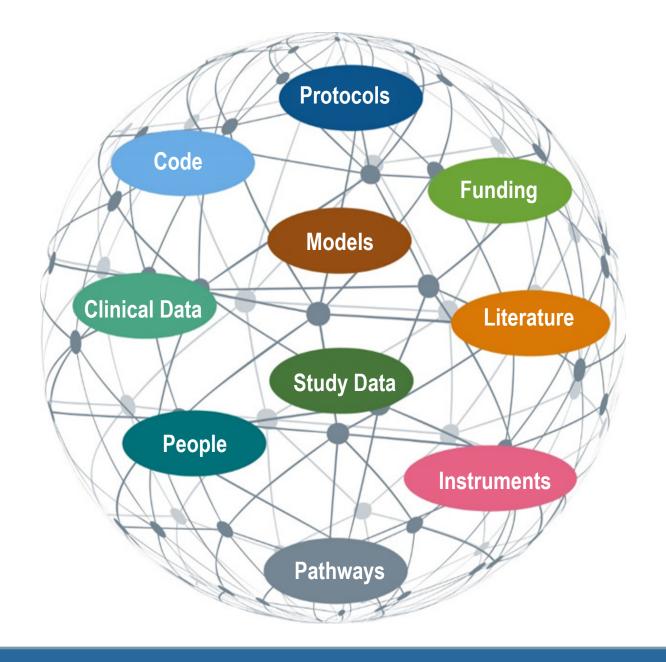
Build a workforce for data-driven research and health

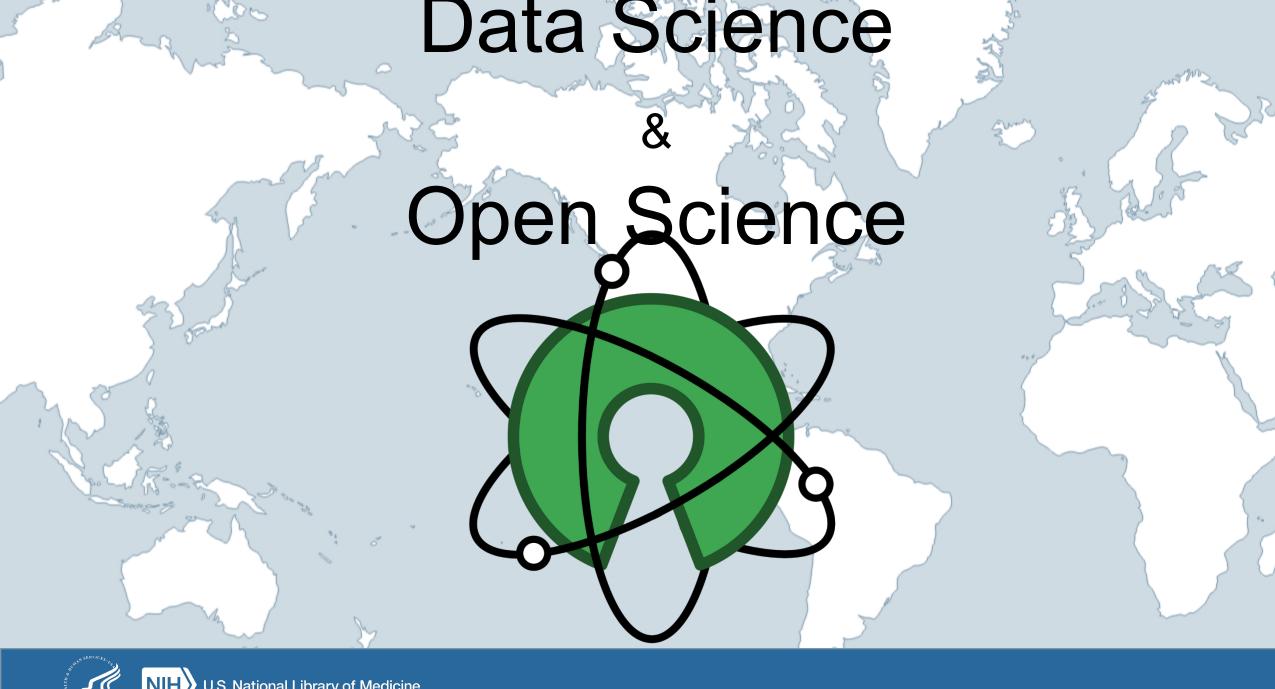




# Fostering a ecosphere of discovery

digital research objects

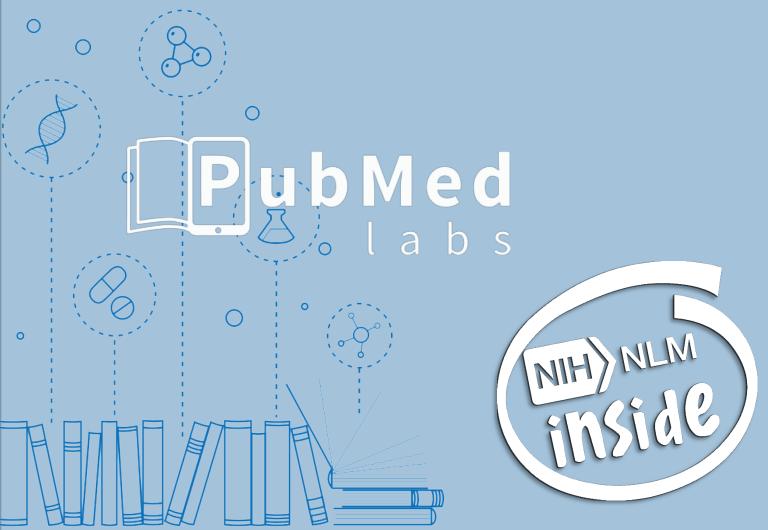






Reach more people in more ways through enhanced dissemination and engagement

Enhance information delivery





## New users, New ways

Biomedical & health information access methods & information dissemination strategies





## The Human Enterprise: Reaching People Wherever







# Build a workforce for data-driven research and health

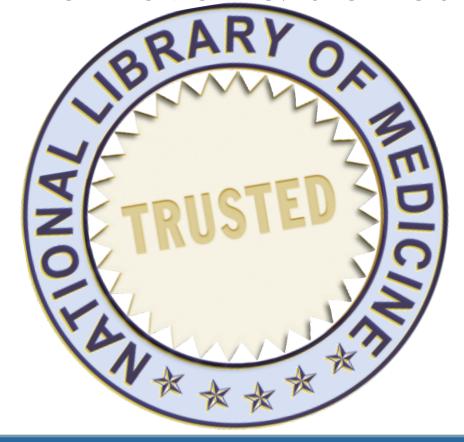




#### Foster distinctiveness of NLM as a

## reliable, trustable source

of health information & biomedical data





#### **NIH Strategic Plan for Data Science**

• Requested by Congress: In its fiscal year 2017 report, the Committee directed NIH to work with its external partners and stakeholders to provide the Committees on Appropriations of the House of Representatives and the Senate a detailed strategic plan by May 5, 2018, spelling out it how intends to make big data sustainable, interoperable, accessible, and usable.

#### Data Infrastructure

GOALS

**OBJECTIVES** 

Optimize data storage and security

 Connect NIH data systems

#### Modernized Data Ecosystem

- Modernize data repository ecosystem
- Support storage and sharing of individual datasets
- Better integrate clinical and observational data into biomedical data science

#### Data Mgmt, Analytics, & Tools

- Support useful, generalizable, & accessible tools and workflows
- Broaden utility of and access to specialized tools
- Improve discovery & cataloging resources

#### Workforce Development

- Enhance the NIH datascience workforce
- Expand the national research workforce
- Engage a broader community

## Stewardship and Sustainability

- Develop policies for a FAIR data ecosystem
- Enhance stewardship



#### NIH Strategic Plan for Data Science:

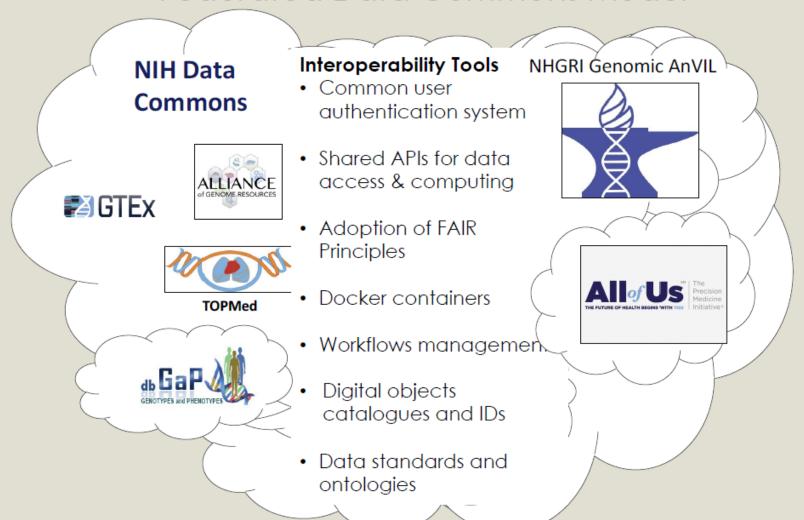
#### **Cross-Cutting Themes**

- Support common infrastructure and architecture on which more specialized platforms can be built and interconnected
- Leverage commercial tools, technologies, services, and expertise; adopt and adapt tools and technologies from other fields
- Enhance the nation's biomedical data-science research workforce: improved training programs, novel
  partnerships
- Ensure NIH-supported data resources are FAIR
- Ensure information security of patient and participant data
- Improve the ability to capture, curate, validate, store, and analyze clinical data for biomedical research
- With community input, develop, promote—and refine as needed—data standards
- Coordinate and collaborate with federal, private and international funding agencies and organizations to promote economies of scale and synergies and prevent unnecessary duplication



## Create a sustainable infrastructure

Federated Data Commons Model



# NIH strongly encourages open access Data Sharing Repositories as a first choice.

https://www.nlm.nih.gov/NIHbmic/nih\_data\_sharing\_repositories.html

#### **Options for scaled implementation for orphan datasets**

Datasets up to 2 gigabytes

#### **PubMed Central**

- PMC stores publicationrelated supplemental materials and datasets directly associated publications. Up to 2 GB.
- Generate Unique Identifiers for the stored supplementary materials and datasets.

Datasets up to **20\*gigabytes** 

### **Dryad and FigShare,** or other

- Assign Unique Identifiers (DOI) to datasets associated with publications and link to PubMed
- Store and manage datasets associated with publication, up to 20\* GB.

High priority Datasets **T-Pbyte range** 

## NIH managed repositoies

- Store and manage large scale, high priority NIH datasets (Partnership with STRIDES)
- Assign Unique Identifiers (dbGap), implement authentication, authorization & access control (eRA Commons)



The 21st Century Collection

- Innovative attribution
- Automated indexing
- Personalized presentation
   & delivery

CUSTODIAN | CONNECTOR | DISCOVERER



## Creating the 21<sup>st</sup> Century Collection of BIOMEDICAL KNOWLEDGE

- Permanent, discoverable archive of text and data
- Listening to trends in Science & Scientific Communication
  - Open Science
  - Preprints & other interim products of research
  - Data Sharing
- Preparing for (and leading) new directions in scientific communications through collaboration



#### **National Library of Medicine Collection**

- ~17,000 serial titles
- All Serials journals, annuals, statistics, etc.
- Discoverable in NLM Catalog & LocatorPlus
- NLM provides ILL and ensures archiving

#### **PubMed**

- 28 M records
- Only bibliographic records

#### MEDLINE

- ~5600 selected journal titles
- Only bibliographic records
- 93% of PubMed

#### **PMC = PubMed Central**

#### Archive

- ~2,000 full participation journals
- 5 million full text articles
- ~2.8 M federally funded public access articles
- Bibliographic records display in PubMed

## NIH Public Access Policy

The NIH Public Access Policy implements Division F Section 217 of PL 111-8 (Omnibus Appropriations Act, 2009). The law states:

The Director of the National Institutes of Health ("NIH") shall require in the current fiscal year and thereafter that all investigators funded by the NIH submit or have submitted for them to the National Library of Medicine's PubMed Central an electronic version of their final, peer-reviewed manuscripts upon acceptance for publication, to be made publicly available no later than 12 months after the official date of publication: Provided, that the NIH shall implement the public access policy in a manner consistent with copyright law.

## Preprints

#### Reporting Preprints and Other Interim Research Products

Notice Number: NOT-OD-17-050

#### **Key Dates**

Release Date: March 24, 2017

Effective date for application: Applications submitted for the May 25, 2017 due date and thereafter

Effective date Research Performance Progress Report (RPPR): RPPRs submitted on or after May 25, 2017

#### Related Announcements

NOT-OD-17-006- Request for Information (RFI): Including Preprints and Interim Research Products in NIH Applications and Reports

#### Issued by

National Institutes of Health (NIH)

#### **Purpose**

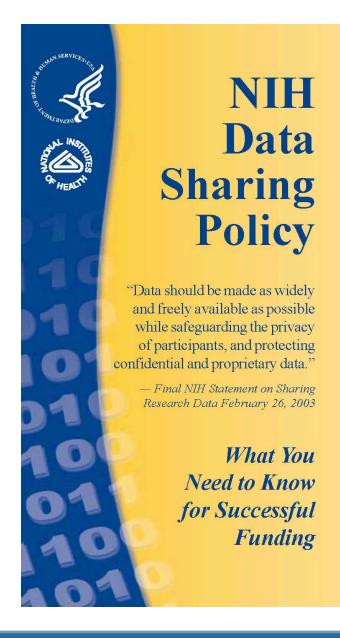
The NIH encourages investigators to use interim research products, such as preprints, to speed the dissemination and enhance the rigor of their work. This notice clarifies reporting instructions to allow investigators to cite their interim research products and claim them as products of NIH funding.

#### **Definitions**

Interim Research Products are complete, public research products that are not final.

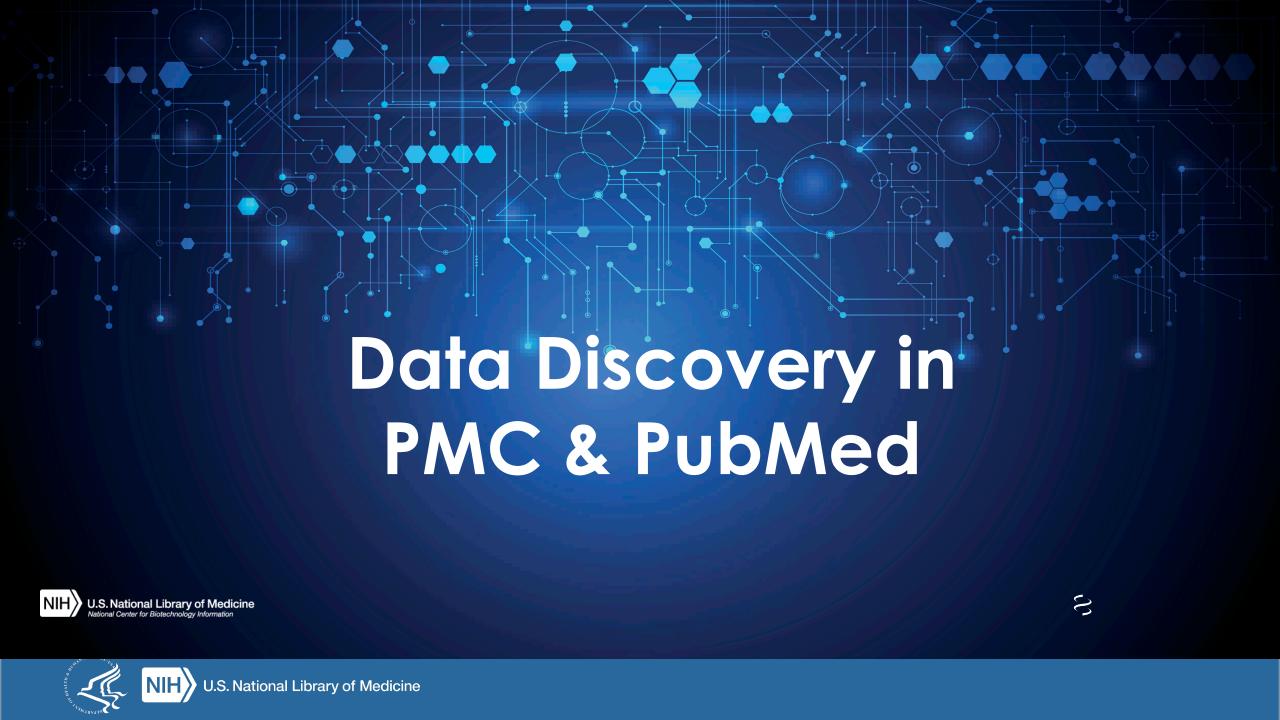
A common form is the preprint, which is a complete and public draft of a scientific document. Preprints are typically unreviewed manuscripts written in the style of a peer-reviewed journal article. Scientists issue preprints to speed dissemination, establish priority, obtain feedback, and offset publication bias.

Another common type of interim product is a preregistered protocol, where a scientist publicly declares key elements of their research protocol in advance. Preregistration can help scientists enhance the rigor of their work.



## NIH Data Sharing Policy

- Data sharing plan in proposals (unscored)
  - Publishing
  - Researcher Efforts
  - Data Enclave
  - Data Archive
  - Mixed Mode
- NIH funds can be used to support data sharing



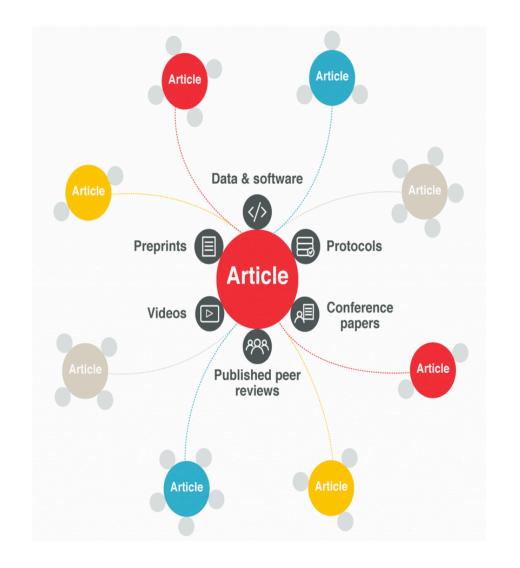
## Journal Article as a Nexus for Discovery

#### PubMed Central (PMC)

contains full text records of publisher's versions and author manuscripts

#### PubMed

contains citations, abstracts, and associated metadata



 $\underline{https://www.crossref.org/blog/the-article-nexus-linking-publications-to-associated-research-outputs/linking-publication-outputs/linking-publication-output$ 

#### NLM Strategic Plan, Objective 1.1

"NLM will stimulate new forms of scientific communication and become the library of the future, one of connections between and among literature, data, models, and analytical tools. Creating efficient ways to link the literature with associated datasets enables knowledge generation and discovery."

# Why link data?

Leverage existing NCBI resources to provide users with new ways to deposit, find, access, and re-use (when applicable) data.

Support <u>NIH efforts</u> to enhance rigor and further support research that is reproducible, robust, and transparent.

## Data Availability Statements in PubMed Central (PMC)

#### **Data Availability**

All relevant data are within the paper and its Supporting Information files. Protein structure files were obtained from Protein Data Bank. Protein structure accession and site mapping are listed in S1 Table (https://doi.org/10.6084/m9.figshare.4910141.v2).

#### Data availabil

The authors declare that all the data supporting the findings of this study are available within the paper and its supplementary information files. Full gels and Western blots are presented in Supplementary Fig. 12. The microarray data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GSE112776 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112776].

#### Data availability

The datasets and computer code produced in this study are available in the following databases: Hi-C and RNA-seq data: Bioproject PRJNA464299 (https://www.ncbi.nlm.nih.gov/bioproject /?term=PRJNA464299).

The fasta sequences of the Syn-HiC region and its native counterpart are available as a supplementary file (Dataset EV1).

Previously published datasets used in this work are described in Table 3.

#### **Examples of what you will find in these sections:**

- "Data available upon request"
- References to supplementary material
- NCBI database accession

#### Data availability

The datasets generated during the current study are available from the corresponding author on reason request.

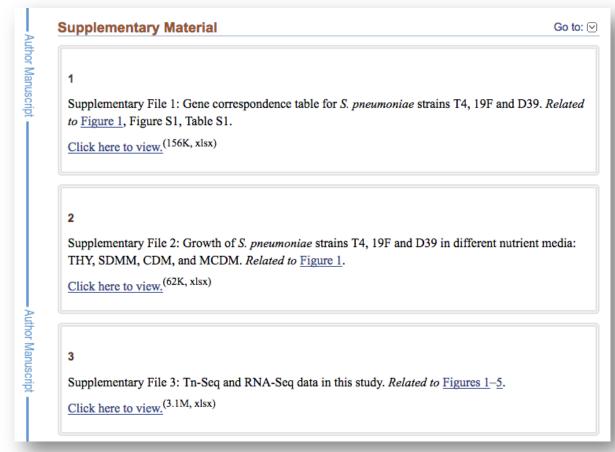
numbers (GEO, GenBank,

dbGap, SRA)

- General data repository record DOIs (figshare, Dryad)
  PMC Search: has data avail[filter]
- Government data repository



# Supplementary Data in PubMed Central (PMC)



PMC Search: <a href="https://has.suppdatafilter">has.suppdatafilter</a>]

Examples of what you will find in supplementary material:

- Computer code
- Mathematical or computational models
- Audio or video clips
- Oversized tables
- Intervention protocols
- Primary or supplementary data sets
- Expanded methodology sections
- Additional figures

Source: Publication Manual of APA (6th ed.)

# Data Citations in PMC

"Data citations should facilitate access to the data and any associated metadata, code, or related materials."

PMC Search: <a href="https://has.data.citations[filter">has.data.citations[filter]</a>

Data Citations

Go to: 

1. Ibanez J. M.. Australian Antarctic Data Centre. 2017 <a href="https://doi.org/10.4225/15/58e6edfc522cf">https://doi.org/10.4225/15/58e6edfc522cf</a>
2. Ibanez J. M.. Australian Antarctic Data Centre. 2017 <a href="https://doi.org/10.4225/15/58e47c1be64b8">https://doi.org/10.4225/15/58e47c1be64b8</a>
3. Ibanez J. M.. Australian Antarctic Data Centre. 2017 <a href="https://doi.org/10.4225/15/58e47c33f2500">https://doi.org/10.4225/15/58e47c33f2500</a>
4. Ibanez J. M.. Australian Antarctic Data Centre. 2017 <a href="https://doi.org/10.4225/15/58e580cfe8c95">https://doi.org/10.4225/15/58e47be626ee5</a>
5. Ibanez J. M.. Australian Antarctic Data Centre. 2017 <a href="https://doi.org/10.4225/15/58e47be626ee5">https://doi.org/10.4225/15/58e47be626ee5</a>

Ibanez J. M., Australian Antarctic Data Centre, 2017 https://doi.org/10.4225/15/58e47c03e652a

```
S Gavrilov, RP Harvey & VE Papaioannou. Mus musculus T-box 2 (Tbx2), mRNA. NCBI Nucleotide. 2014:NM_009324.2. NM_009324
        <!--Data reference: NCBI Nucleotide: pub-id-type="accession" assigning-authority="NCBI"
        contains version details -->
        <ref id="bib13">
          <element-citation publication-type="data">
            <person-group person-group-type="author">
                <surname>Gavrilov</surname>
                <given-names>S</given-names>
              </name>
                <surname>Harvey</surname>
                <given-names>RP</given-names>
                <surname>Papaioannou</surname>
                <given-names>VE</given-names>
              </name>
            </person-group>
            <year iso-8601-date="2014">2014
            <data-title>Mus musculus T-box 2 (Tbx2), mRNA</data-title>
            <source>NCBI Nucleotide</source>
            <pub-id pub-id-type="accession" assigning-authority="NCBI"</pre>
              xlink:href="http://www.ncbi.nlm.nih.gov/nuccore/120407038"
              >NM 009324</pub-id>
            <version designator="NM_009324.2">NM_009324.2/version>
          </element-citation>
        </ref>
```

https://www.ncbi.nlm.nih.gov/pmc/pmcdoc/tagging-guidelines/article/dobs.html#dob-datacitationshttp://iats4r.org/data-citations



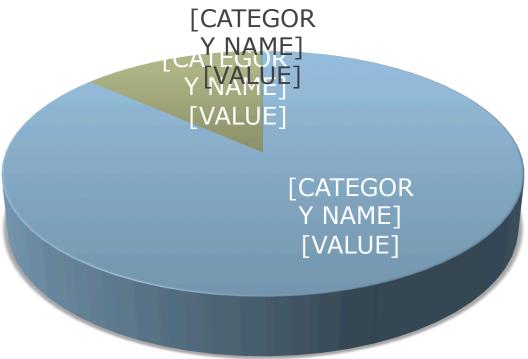


# Current snapshot of data in PMC

~20% of articles in PMC (~930,000) have associated supplementary material, a data availability statement, and/or data citation(s)

#### **Challenge:**

Inconsistent (or non-existent) policies on how to make data available and cite it at the funder and journal level result in inconsistent data association practices



Source: PubMed Central, 16 July 2018

### Data Links in PubMed: **Secondary Source ID**

#### Secondary Source ID data sources:

- Publishers
- NLM indexers
- PMC

#### Challenges:

- NLM indexing resources were reduced in 2016, leading NLM to explore alternative options
- No incentives for journals/ publishers to supply these metadata to NLM

Andrologia, 2017 Oct;49(8), doi: 10.1111/and.12712, Epub 2016 Oct 26.

#### Characterisation of pomegranate juice effects on human corpus cavernosum.

Gur S<sup>1,2</sup>, Rezk BM<sup>3</sup>, Abd Elmageed ZY<sup>2</sup>, Kadowitz PJ<sup>2</sup>, Sikka SC<sup>2</sup>, Hellstrom WJG<sup>2</sup>.

Author information

Pomegranate (POM) juice may benefit the erectile process, but the scientific evidence is lacking. This study evaluates the molecular characterisation and confirmation of POM's action on human corpus cavernosum (HCC) obtained from patients (n = 16) undergoing penile prosthesis implantation. After phenylephrine contraction, the relaxant effects of POM with various inhibitors in the presence and absence of palmitic acid (PA)-induced acute oxidative stress were investigated. Electrical field stimulation (EFS)- and acetylcholine (ACh)-induced relaxation were performed using organ bath preparation. Expression of neuronal nitric oxide synthase (nNOS), endothelial (eNOS), phosphodiesterase (PDE)-5A and cGMP levels were assessed in cells from ex vivo organ cultures of HCC, using RT-PCR, ELISA and immunohistochemistry techniques. POM induced marked relaxation of HCC (maximum response: 97.0 ± 3.1%) and reversed the PA-induced decrease of EFS (20 Hz), nNOS transcription was increased by 7-fold in POM-treated cells without influencing eNOS and PDE5A expressions. We conclude that POM induced marked relaxation of HCC via: (i) nNOS stimulation, and (ii) downstream relaxation stimulated by nNOS and cGMP and bypassing the NO and PDE5. This action provides a rationale for the therapeutic or preventative use of POM in men with erectile dysfunction who do not respond well to PDE5 inhibitors.

© 2016 Blackwell Verlag GmbH.

KEYWORDS: eNOS and PDE5A; erectile dysfunction; human corpus cavernosum; nitric oxide; pomegranate

PMID: 27785815 DOI: 10.1111/and.12712







Secondary source ID

Secondary source ID GENBANK/NM 000620 GENBANK/NM 000603 GENBANK/NM 033437 **GENBANK/NM** 001101.2

Example PubMed Searches: <a href="https://hasdatabanklist">hasdatabanklist</a> genbank[si] OR figshare[si]



Ann Med. 2017 Dec;49(8):718-727. doi: 10.1080/07853890.2017.1372622. Epub 2017 Sep 6.

Baseline fatty acids, food groups, a diet score and 50-year all-cause mortality rates. An ecological analysis of the Seven Countries Study.

Menotti A1, Kromhout D2,3,4, Puddu PE5, Alberti-Fidanza A6, Hollman P3, Kafatos A7, Tolonen H8, Adachi H9, Jacobs DR Jr4.

#### Abstract

OBJECTIVES: This analysis deals with the ecologic relationships of dietary fatty acids, food groups and the Mediterranean Adequacy Index (MAI, derived from 15 food groups) with 50-year all-cause mortality rates in 16 cohorts of the Seven Countries Study.

MATERIAL AND METHODS: A dietary survey was conducted at baseline in cohorts subsamples including chemical analysis of food samples representing average consumptions. Ecologic correlations of dietary variables were computed across cohorts with 50-year all-cause mortality

RESULTS: There was a 12-year average age at death population difference between extreme cohorts. In the 1960s the average population intake of saturated (S) and trans (T) fatty acids and hard fats was high in the northern European cohorts while monounsaturated (M), polyunsaturated (P) fatty acids and vegetable oils were high in the Mediterranean areas and total fat was low in Japan. The 50-year all-cause mortality rates correlated (r=-0.51 to -0.64) ecologically inversely with the ratios M/S, (M+P)/(S+T) and vegetable foods and the ratio hard fats/vegetable oils. Adjustment for high socio-economic status strengthened (r= -0.62 to -0.77) these associations including MAI diet score.

CONCLUSION: The protective fatty acids and vegetable oils are indicators of the low risk traditional Mediterranean style diets. KEY MESSAGES We aimed at studying the ecologic relationships of dietary fatty acids, food groups and the Mediterranean Adequacy Index (MAI, derived from 15 food groups) with 50-year all-cause mortality rates in the Seven Countries Study. The 50-year all-cause mortality rates correlated (r = -0.51 to -0.64) ecologically inversely with the ratios M/S [monounsaturated (M) + polyunsaturated (P)]/[saturated (S) + trans (T)] fatty acids and vegetable foods and the ratio hard fats/vegetable oils. After adjustment for high socio-economic status, associations with the ratios strengthened (r = -0.62 to -0.77) including also the MAI diet score. The protective fatty acids and vegetable oils are indicators of the low risk traditional Mediterranean style diets.

KEYWORDS: 50-year follow-up; Seven Countries Study; all-cause mortality; fatty acids; socio-economic status

PMID: 28847158 DOI: 10.1080/07853890.2017.1372622





LinkOut - more resources

**Full Text Sources** Taylor & Francis

#### Other Literature Sources

Supporting Data and Materials for the article - figshare

loprovfigshare[Filter] OR loprovdryaddb[Filter]

### **Data Links in PubMed: LinkOut**

Links to the materials directly supporting the research discussed in the cited article, including data sets from experiments/studies accessory graphics, images, sound, and multimedia files related to the article.

American Society for Microbiology - videos and podcasts (asmmedia)

DataMed.org (DataMed)

Dryad Digital Repository (dryaddb)

figshare (figshare)

NCI caNanoLab Data Portal (ncicananolab)

NHGRI Bioinformatics and Scientific Programming Core (nhgribspc)

NIMH Data Archive (NDA)

Nuclear Receptor Signaling Atlas (NURSA) (nursa)

Princeton University MicroArray database (PUMAdb)

Surface Management System Database (SumsDB)



Event-related Potential Data From a Guess the

Number Brain-Computer Interface Experiment on School Children

R Mouček, L Vařeka ... P Brůha + expand PubMed: 28350376 PMC: PMC5369320

Sci Data, 4, 160121 © 2017 Mar 28

#### Abstract

Guess the number is a simple P300-based brain-computer interface experiment. Its aim is to ask the measured participant to pick a number between 1 and 9. Then, he or she is exposed to corresponding visual stimuli and experimenters try to guess the number thought while they are observing event-related potential waveforms on-line. 250 school-age children participated in the experiments that were carried out in elementary and secondary schools in the Czech Republic. Electroencephalographic data from three EEG channels (Fz, Cz, Pz) and stimuli markers were stored. Additional metadata about the participants were collected (gender, age, laterality, the number thought by the participant, the guess of the experimenters, and various interesting additional information). Consequently, we offer the largest publicly available odd-ball paradigm collection of datasets to neuroscientific and brain-computer interface community.

DOI: 10.1038/sdata.2016.121

#### Conflict of interest statement

The authors declare no competing financial interests.



# Future Direction: Data Citations in PubMed Labs



#### References

#### **Data Citations**

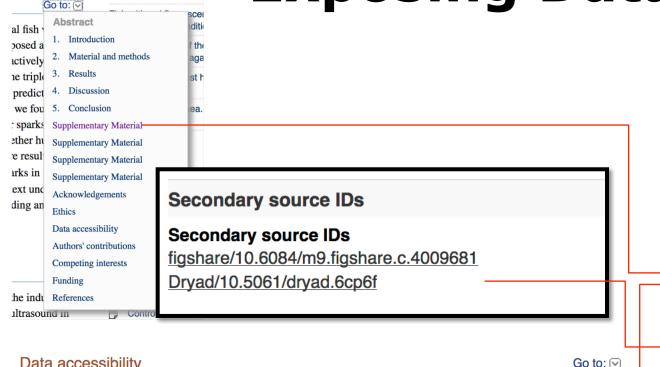
- Mouček R., Vařeka L., Prokop T., Štebeták J., Brůha P.. EEG/ERP Portal (EEGbase), Experiments package: PROJECT DAYS P3 NUMBERS. 2016 https://eegdatabase.kiv.zcu.cz
- Mouček R., Vařeka L., Prokop T., Štebeták J., Brůha P.. Harvard Dataverse. 2016 http://dx.doi.org/10.7910/DVN/TDURF9 - DOI
- Mouček R., Vařeka L., Prokop T., Štebeták J., Brůha P., Harvard Dataverse. 2016 http://dx.doi.org/10.7910/DVN/NJWBL6 - DOI
- Mouček R., Vařeka L., Prokop T., Štebeták J., Brůha P.. Harvard Dataverse. 2016 http://dx.doi.org/10.7910/DVN/6RKLUX - DOI
- Mouček R., Vařeka L., Prokop T., Štebeták J., Brůha P.. Harvard Dataverse. 2016 http://dx.doi.org/10.7910/DVN/OZRAIV - DOI

https://www.ncbi.nlm.nih.gov/labs/pubmed/28350376/





### **Future Direction: Exposing Data in PMC**



Data accessibility

deiaisi</i>

Data for spectral properties of ocular sparks, copepod eye reflectance and all data of the two experiments are available at http://dx.doi.org/10.5061/dryad.6cp6f [48].

48. Michiels NK, Seeburger VC, Kalb N, Meadows MG, Anthes N, Mailli AA, Jack CB. 2018. Data from: Controlled iris radiance in a diurnal fish looking at prey Dryad Digital Repository (doi:10.5061/dryad.6cp6f)



Published online 2018 Jul 10. doi: 10.1038/sdata.2018.131 Data Descriptor

PMCID: PMC6038849 PMID: 29989595

Transcriptome sequencing and molecular markers discovery in the gonads of Portunus sanguinolentus

Yin Zhang, 1,2 Guidong Miao, 1,2 Qingyang Wu, 1,2 Fan Lin, 1,2 Cuihong You, 1,2 Shuqi Wang, 1,2 Jude Juventus Aweya, 1,2 and Hongyu Maa,1,2

Author information ▶ Article notes ▶ Copyright and License information ▶ Disclaimer



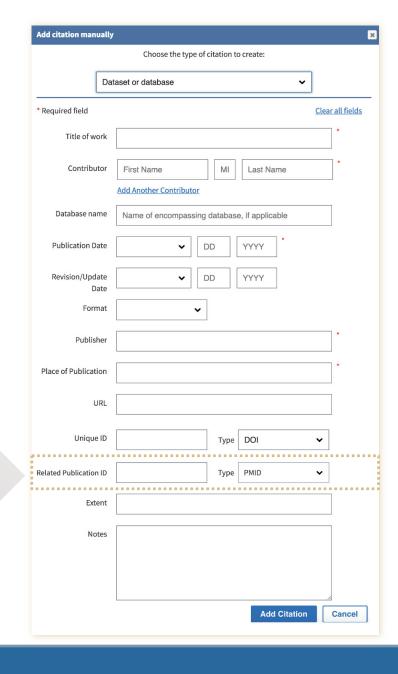
figshare/10.6084/m9.figshare.c.3982629

Abstract

Go to: ₩

Crab culture has gained prominence in the last decade due to the large global market demand for live crabs and crab products. Portunus sanguinolentus is one of the economically important crab species in the Indo-Pacific region, with distinct differences in growth and size between male and female crabs, thus, leading to huge difference in their market values. The culture of P. sanguinolentus is still in its infancy, with crab

# Future Direction: Linking Publications to Data in My Bibliography







### ClinicalTrials.gov is a database of privately and publicly funded clinical studies conducted around the world.

#### Explore 279,151 research studies in all 50 states and in 204 countries.

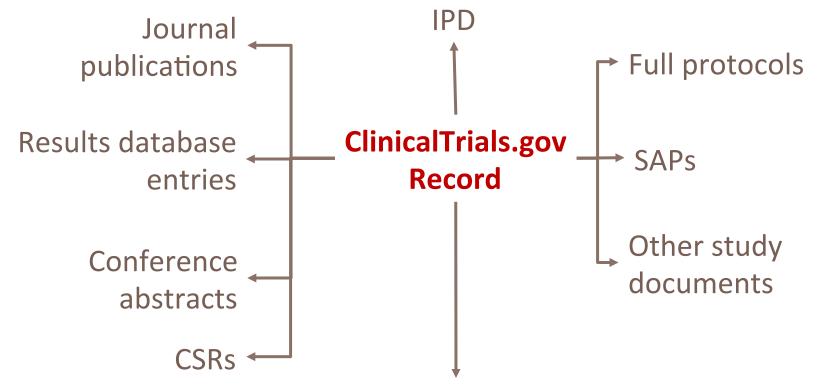
ClinicalTrials.gov is a resource provided by the U.S. National Library of Medicine.

**IMPORTANT**: Listing a study does not mean it has been evaluated by the U.S. Federal Government. Read our disclaimer for details.

Before participating in a study, talk to your health care provider and learn about the <u>risks and</u> potential benefits.

<ul><li>Recruiting</li><li>All studies</li></ul>	and not yet recruiting	studies			
Condition or dis	ease () (For example: bre	east cancer)			
				×	
Other terms ① (F	or example: NCT number,	drug name, ir	vestigator nam	ie)	
				x	
Country <b>1</b>				x	
Country (1)				x	
Country (1)				<b>x * x</b>	

# ClinicalTrials.GOV: Informational Scaffold



Other Information (e.g., press releases, news articles, editorials)



### Common Data **Elements**



Search Common Data Elements

**Q** Search

The NIH Common Data Elements (CDE) Repository has been designed to provide access to structured human and machine-readable definitions of data elements that have been recommended or required by NIH Institutes and Centers and other organizations for use in research and for other purposes. Visit the NIH CDE Resource Portal for contextual information about the repository.

The Repository is a platform for identifying related data elements in use across diverse areas, for harmonizing data elements, and for linking CDEs to other existing standards and terminologies, including the value sets in the Value Set Authority Center (VSAC).

#### Search

Search for individual data elements, by definition, users or sources. Search for sets of data elements ("boards") identified by a particular group for a particular use (e.g. particular research solicitation).

#### Compare / Harmonize

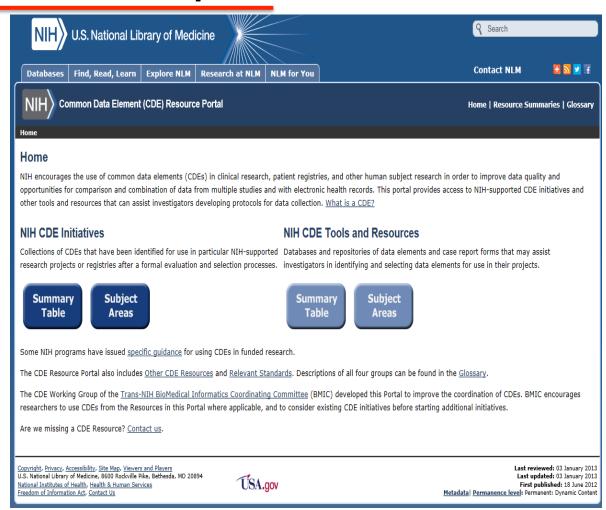
Analyze and resolve differences between data elements. Assure that your forms are using variables that will be usable by certified EHRs.

#### Create

Draw upon the experience of colleagues and others to design unique data elements and measures.

# Common Data Elements Make data findable, interoperable

- Structured human & machine readable definitions of NIH CDEs allowing
  - ◆Search for individual CDE or sets per programs
  - ◆Compare & harmonize similar but distinct CDEs
  - Select or create CDEs with minimal duplication



Explore NLM

Soarch:









Trans-NIH BioMedical Informatics Coordinating Committee (BMIC)

Research at NLM

**NLM for You** 

**BMIC Home** 

**CDE Resource Portal** 

Home > BMIC Home

Show 50 Aptries

**Databases** 

#### NIH Data Sharing Repositories

Find, Read, Learn

This table lists NIH-supported data repositories that make data accessible for reuse. Most accept submissions of appropriate data from NIH-funded investigators (and others), but some restrict data submission to only those researchers involved in a specific research network. Also included are resources that serve as a portal for information about biomedical data and information sharing systems. The table can be sorted by repository name and by NIH Institute or Center and may be searched using keywords so that you can find repositories more relevant to your data. Links are provided to information about submitting data to and accessing data from the listed repositories. Also, we have asked all repositories to indicate whether or not they have—current NIH funding support; open data submission; open data access; open time frame for data submission; and sustained support. Their Yes or No responses are included in the five columns to the right of the table. The information in the BMIC list is also available in a downloadable Excel version.

In our guery to the repositories, we acknowledged that some repositories may have restrictions or limitations on data submission and requirements for data access. Additional information about the repositories and points-of-contact for further information or inquiries can be found on the websites or links shown for each. Are we missing a data sharing repository that should be included in this list? Contact us.

Search: Search:										
	ICO ^	Repository Name 🍦	Repository Description	Data Submission Policy	Access to Data	Current NIH funding support	Open data submission	Open data access	Open time frame for data deposit	Sustained support
	Common Fund	<u>Epigenomics</u>	Epigenomic, 6 histone modification marks, DNAse I, DNA methylation, transcriptome for wide variety of cell types and tissues.	Not applicable	How to access Epigenomics data	No	No	Yes	No	Yes
	Common Fund	exRNA Atlas	Includes exRNA profiles derived from various biofluids and conditions and currently stores data profiled from small RNA sequencing assays.	Not applicable	How to access exRNA Atlas data	Yes	No	Yes	No	Yes
	Common Fund	GTEx	The Genotype-Tissue Expression (GTEx) project aims to provide to the scientific community a resource with which to study human gene expression and regulation and its	Project data only	How to access GTEx data	Yes	No	Yes	No	Yes



# Office of Data Science Strategy NIH Office of the Director

Susan Gregueric, PhD
Senior Advisor to the Director for Data
Science

### Data Science Fellowships @ NIH

Jessica Mazerik and Susan Gregurick







# NIH Data Science Fellowships – Happening NOW

#### **Graduate Data Science Summer Program**

- OITE in partnership with ODSS
- Consortium of local universities, but students from all universities can apply
  - UVA, George Mason, George Washington, UMD
- Pilot driven by discussion with local universities consortium
- Masters students for summer program at NIH
- Intramural placement
- ~15 interns for summer 2019
- Call for projects late November



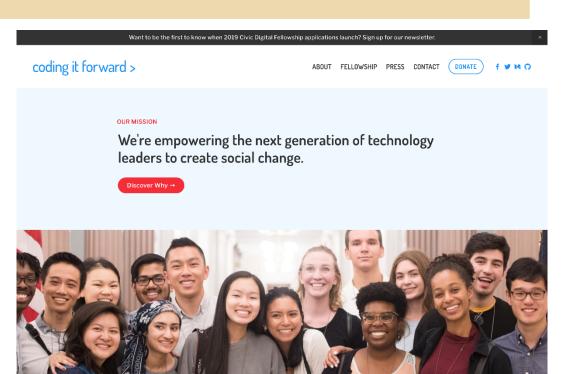
https://www.training.nih.gov/data\_science



# NIH Data Science Fellowships – Happening NOW

#### **Coding it Forward**

- Civic Digital Fellowship program
- Student-led non-profit to place tech-savvy students in federal agencies; pipeline to public service for technology students
- Mentor/project call for 2019
- 10-12 week summer program
- Placement in administrative/extramural funding offices at NIH
- ODSS will coordinate some central NIH activities so fellows connect with each other on campus
- ~10 fellows for summer 2019





# NIH Data Science Fellowships – Happening SOON

#### NIH Data and Technology Advancement (DATA) Fellows

#### **Overview**

- One or two year positions in high-impact NIH programs
- Data science experts, not necessarily familiar with biomedical research
- Opportunity to work with large volumes of biomedical research data; impact on public health
- Learn from NIAID's program
  - https://www.zintellect.com/Opportunity/Details/NIAID-17-01
- ~2-5 fellows in first cohort

#### PLAN

- Goal: first cohort starts fall 2019
- Launch multiple onboarding/training mechanisms to apply
- Advantages flexibility in salary, no one-size fits all
- Program evaluation in 2024

### **Provisions for Data Management** and

**Sharing Policy** 

Request for Information (RFI) on Proposed Provisions for a Draft Data Management and Sharing Policy for NIH Funded or Supported Research NIH Funded or Supported Research Notice Number: NOT-OD-19-014 Introduction

The purpose of this Notice is to solicit public input on proposed key provisions that could serve as a foundation for a future NIH policy for data management and sharing.

Background Release Date: October 10, 2018 Response Date: December 10, 2018 Key Dates Background

NIH has a longstanding commitment to making the results and accomplishments of the research that it funds and confidential and proprietary data. Increasing access to scientific data made as widely and freely available as possible while safeguarding the privacy of participants and professing accomplishments. Related Announcements NIH has a longstanding commitment to making the results and accomplishments of the research that it funds and conducts available to the public. In NIH's view, data should be as widely and freely available as possible while safeguarding the privacy of participants and protecting confidential and properties. Specifically, sharing of scientific and accomplishments of the research that it funds and conducts available to the public. In NIH's view, data should be made as widely and freely available as possible while safeguarding the privacy of participants and protecting stewardship, over taxpayer funds. Specifically, sharing of scientific and reflects NIH's responsibility to maintain stewardship, over taxpayer funds. The service of the second of t resulting from NIH funding for support offers many benefits and reflects the fundings, on investment of scientific research funding.

The sum of the sum o results enables researches to more vigorously test the validity of research findings, and spraying the analyses by combining appropriate identification, adoption, and crediting of grounds are search findings, and appropriate identification, adoption, and crediting of grounds are search findings, and appropriate identification, adoption, and provided in the research findings, and appropriate identification, adoption, and provided in the research findings, and appropriate identification, adoption, and provided in the research findings, and procedures to improve health and prevent disease. Effective data sharing also informs future research pathways, increases the return on investment of procedures and procedures to improve health and prevent disease. NOT-OD-17-015 NOT-OD-18-134 Trontiers. Data sharing and procedures to improve health and prevent disease. Effective With the FAIR (Findable, Accessible, Interroperable, and Sharing Practices, thus, NIH encourages data sharing consistent with the FAIR (Findable, Accessible, Interroperable, and Sharing Practices, thus, NIH encourages data sharing consistent with the FAIR (Findable, Accessible, Interroperable, and Practices, thus, NIH encourages data sharing consistent with the FAIR (Findable, Accessible, Interroperable, and Practices, thus, NIH encourages data sharing consistent with the FAIR (Findable, Accessible, Interroperable, and Practices, thus, NIH encourages data sharing consistent with the FAIR (Findable, Accessible, Interroperable, Interroperab NOT-OD-19-034 Nonwedge, products, and procedures to improve health and prevent data sharing on signific networks that with the FAIR (Findable, Fundable, Fundabl National Institutes of Health (NIH) data management and sharing practices, thus, NIH encourages data sharing consistent with the FAIR (Findable, Accessible, Interoperable, and Re-usable) and principles. To help NIH plan). To help NIH plan) and Digital Scientific Data from NIH Funded Scientific Research (NIH Plan). And, in 2017, one of the principle of the princip Issued by In 2015, NIH released its Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Data Management, Sharing, and NIH's thinking about key provide priorities for data management and sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for Information on Strategies for NIH Data Management, Sharing, NIH in 2016 released a Request for NIH Data Management, Sharing, NIH in 2016 released a Request for NIH Data Management, Sharing, NIH in 2016 released a Request for NIH Data Management, Sharing, NIH in 2016 released a Request for NIH Data Management, Sharing, NIH in 2016 released a Request for NIH Dat Priorities for data management and sharing. NH in 2016 released a Request for Information on Strategies for NH Data Sharing Policy.

With the National Science Foundation a joint workshop, focused on the value of data sharing Policy. Purpose



#### What Do We Need to Know?

- Before we can start talking about a NIH policy for sharing and managing data, we need to think about what the key components are, such as:
  - The definition of scientific data
  - The requirements for data management and sharing plans
  - The optimal timing to consider in implementing various parts of a new policy
    - Phased adoption?
    - Incremental implementation?

# Ongoing Implementation Considerations

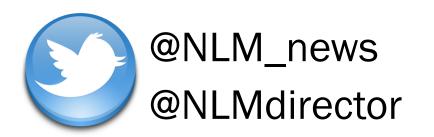
- Budgeting for data management and sharing
- Use of existing repositories; criteria for identifying
- Community expectations for shared data



## Questions and discussion

## Reaching NLM







patti.brennan@nih.gov