

Data Repositories For Research Reproducibility

*Kathy Davies, MLS
Interim Director of Libraries*

*Jennifer Putnam Davis, MA, MLIS
Scholarship and Data Librarian
Greenblatt Library*



AUGUSTA UNIVERSITY
University Libraries

Today's Topics

- Overview of Data Reproducibility
- Data Reproducibility Components and Challenges
- Data Reproducibility Initiatives
- Data Journals and Repositories
- University Library Resources

Within the scope of the health sciences, social sciences, and the arts and humanities disciplines



Data Reproducibility

- Definitions
 - Reproducibility

“Ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator.”

- Replicability

“Ability of a researcher to duplicate the results of a prior study if the same procedures are followed but new data are collected.”

WHAT DOES RESEARCH REPRODUCIBILITY MEAN?

STEVEN N. GOODMAN*, DANIELE FANELLI AND JOHN P. A. IOANNIDIS

Science Translational Medicine 01 Jun 2016:

Vol. 8, Issue 341, pp. 341ps12 DOI: 10.1126/scitranslmed.aaf5027



Data Reproducibility Components and Challenges in the Health Sciences and Social Sciences



Data Reproducibility Components

- Sharing analytical data sets (accessibility)
- Relevant metadata
- Analytical code
- Related software
- Verify data is accurately represented



Why Research Reproducibility

- Evidence that results from experiment/study are objective
- Address possible bias in reported results
- Used in decision-making creating risk of potential harm
 - Patient care
 - Public health
 - Engineering
 - Environmental health
- Avoid reliance on invalid research/data
- NIH grant applications address rigor & reproducibility
 - Old language: *scientific premise*
 - New language: *rigor of prior research* (1/25/19)



NIH ENHANCING REPRODUCIBILITY GUIDELINES

what you need to know

WHAT ARE THE FOUR ELEMENTS OF RIGOR?

1

RIGOR OF
THE PRIOR
RESEARCH

2

RIGOR OF THE
PROPOSED
RESEARCH

3

BIOLOGICAL
VARIABLES

4

AUTHENTICATION

Send inquiries to
reproducibility@nih.gov

See also NIH Notice NOT-OD-18-228
<https://grants.nih.gov/grants/guide/notice-files/NOT-OD-18-228.html>

WHERE IN THE APPLICATION?

1 RESEARCH STRATEGY

The research strategy is where you discuss the significance, innovation, and approach of your research plan. Let's look at an R01, for example:



Introduction to
Resubmission
and Revision
Applications



Specific
Aims



Research
Strategy



Progress
Report
Publication
List



Vertebrate
Animals

The research strategy guidelines require that you:

- Describe the strengths and weaknesses in the rigor of the prior research that serves as key support.
- Describe plans to address weaknesses in the rigor of the prior research.
- Describe how your experimental design and methods will achieve robust and unbiased results.
- Explain how relevant biological variables, such as sex, are factored into research designs and analyses.

2 ATTACHMENT FOR AUTHENTICATION OF KEY BIOLOGICAL AND/OR CHEMICAL RESOURCES

You must briefly describe methods to ensure the identity and validity of key biological and/or chemical resources used in the proposed studies.

These include, but are not limited to:

CELL LINES

ANTIBODIES

SPECIALTY CHEMICALS

OTHER BIOLOGICS

Standard laboratory reagents that are not expected to vary do not need to be included in the plan. Examples are buffers and other common biologicals or chemicals.

- ☒ **DO NOT** put experimental methods or preliminary data in this section
- ☒ **DO** focus on authentication and validation of key resources

3 REVIEW GUIDELINES

Here are the additional criteria the reviewers will be asked to use:

- Is the **prior research** that serves as the key support for the proposed project rigorous?
- Have the investigators included plans to **address weaknesses in the rigor of prior research** that serves as the key support for the proposed project?
- Have the investigators presented **strategies to ensure a robust and unbiased approach**, as appropriate for the work proposed?
- Have the investigators presented adequate plans to address **relevant biological variables**, such as sex, for studies in vertebrate animals or human subjects?



Reviewers will
also be asked to
comment on that
new attachment
(see Update 2)!

<https://grants.nih.gov/policy/reproducibility/guidance.htm>



AUGUSTA UNIVERSITY
University Libraries

Enhancing Reproducibility in NIH Applications: Resource Chart

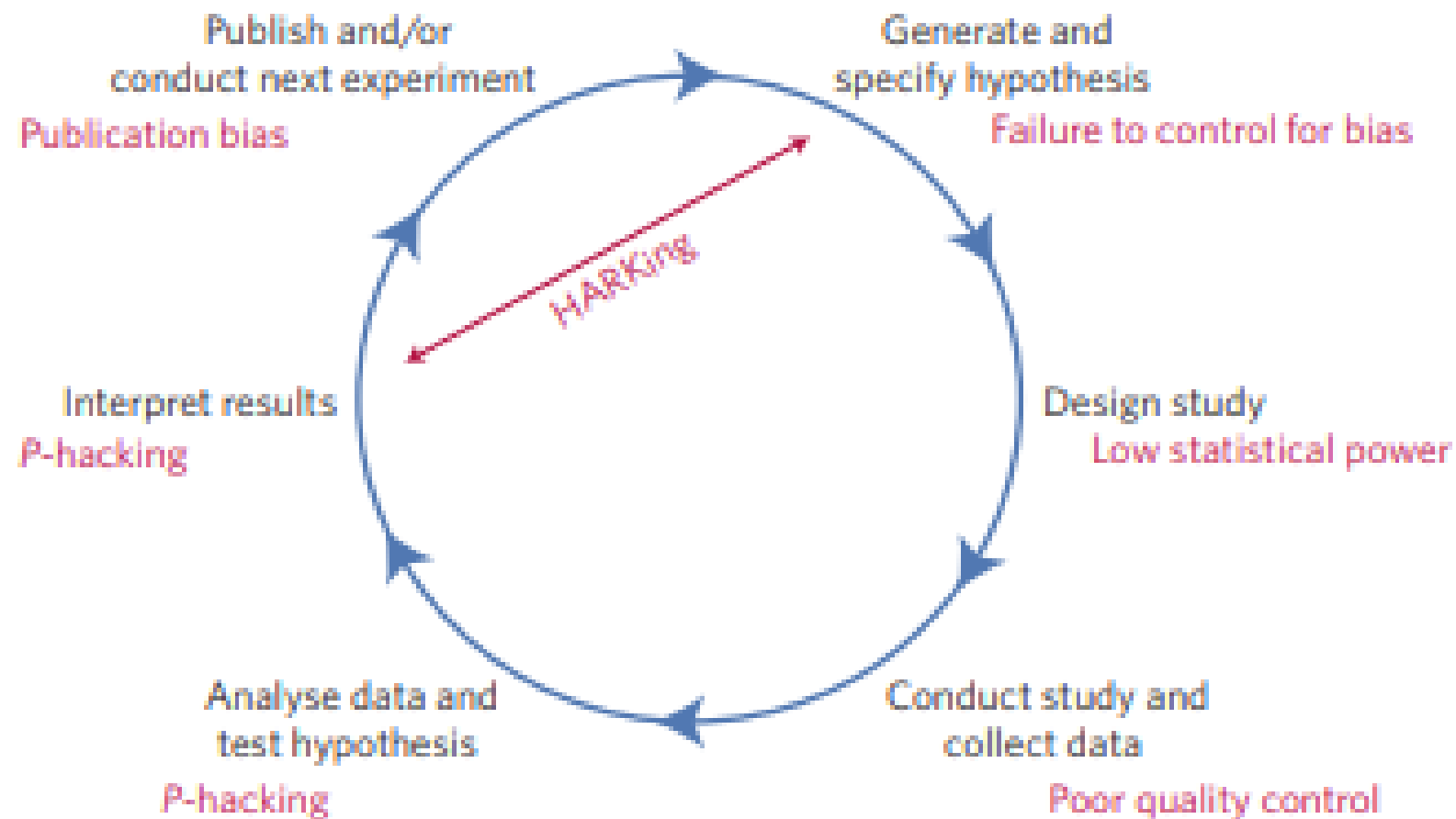
NIH Grants Policy Website: <http://grants.nih.gov/reproducibility/index.htm>

NIH Website: <https://www.nih.gov/research-training/rigor-reproducibility>

4 AREAS OF FOCUS	WHAT DOES IT MEAN?	WHERE SHOULD IT BE INCLUDED IN THE APPLICATION?
Rigor of the Prior Research	<p>A careful assessment of the rigor of the prior research that serves as the key support for a proposed project will help applicants identify any weaknesses or gaps in the line of research.</p> <p>Describe the strengths and weaknesses in the rigor of the prior research (both published and unpublished) that serves as the key support for the proposed project.</p> <p>Describe plans to address weaknesses in the rigor of the prior research that serves as the key support for the proposed project</p> <p><i>*See related FAQs, blog post</i></p>	<p>Research Strategy</p> <ul style="list-style-type: none"> ➤ Significance ➤ Approach
Scientific Rigor (Design)	<p>Scientific rigor is the strict application of the scientific method to ensure robust and unbiased experimental design, methodology, analysis, interpretation and reporting of results.</p> <p>Emphasize how the experimental design and methods proposed will achieve robust and unbiased results.</p> <p><i>*See related FAQs, blog post, examples from pilots</i></p>	<p>Research Strategy</p> <ul style="list-style-type: none"> ➤ Approach
Biological Variables	<p>Biological variables, such as sex, age, weight, and underlying health conditions, are often critical factors affecting health or disease. In particular, sex is a biological variable that is frequently ignored in animal study designs and analyses, leading to an incomplete understanding of potential sex-based differences in basic biological function, disease processes and treatment response.</p> <p>Explain how relevant biological variables, such as the ones noted above, are factored into research designs, analyses, and reporting in vertebrate animal and human studies. Strong justification from the scientific literature, preliminary data or other relevant considerations must be provided for applications proposing to study only one sex.</p> <p><i>*See related FAQs, blog posts, article</i></p>	<p>Research Strategy</p> <ul style="list-style-type: none"> ➤ Approach
Authentication	<p>Key biological and/or chemical resources include, but are not limited to, cell lines, specialty chemicals, antibodies and other biologics.</p> <p>Briefly describe methods to ensure the identity and validity of key biological and/or chemical resources used in the proposed studies. These resources may or may not have been generated with NIH funds and:</p> <ul style="list-style-type: none"> • may differ from laboratory to laboratory or over time; • may have qualities and/or qualifications that could influence the research data; • are integral to the proposed research. <p>The authentication plan should state in one page or less how you will authenticate key resources, including the frequency, as needed for your research. Note: Do not include authentication data in your plan.</p> <p><i>*See related FAQs, blog post, examples</i></p>	<p>Other Research Plan Section</p> <ul style="list-style-type: none"> ➤ Include as an attachment ➤ <u>Do not include</u> in the Research Strategy.



Data Reproducibility Challenges: Research Life Cycle



- P- hacking: making data “fit” desired outcomes

A manifesto for reproducible science

Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware & John P. A. Ioannidis
Nature Human Behaviour volume 1, Article number: 0021 (2017)



Data Reproducibility: Disciplinary Practice

Degree of determinism

Signal to measurement-error ratio

Complexity of designs and measurement tools

Closeness of fit between hypothesis and experimental design or data

Statistical or analytic methods to test hypotheses

Typical heterogeneity of experimental results

Culture of replication, transparency, and cumulating knowledge

Statistical criteria for truth claims

Purposes to which findings will be put and consequences of false conclusions

- Robustness

“stability of experimental conclusions to variations”

- Generalizability

“persistence of effect in settings different from experimental framework”

- Publishing Experimental Design

Level of detail required or typical for different fields or journals

WHAT DOES RESEARCH REPRODUCIBILITY MEAN?

STEVEN N. GOODMAN*, DANIELE FANELLI AND JOHN P. A. IOANNIDIS



Scope of Problem: Health Sciences

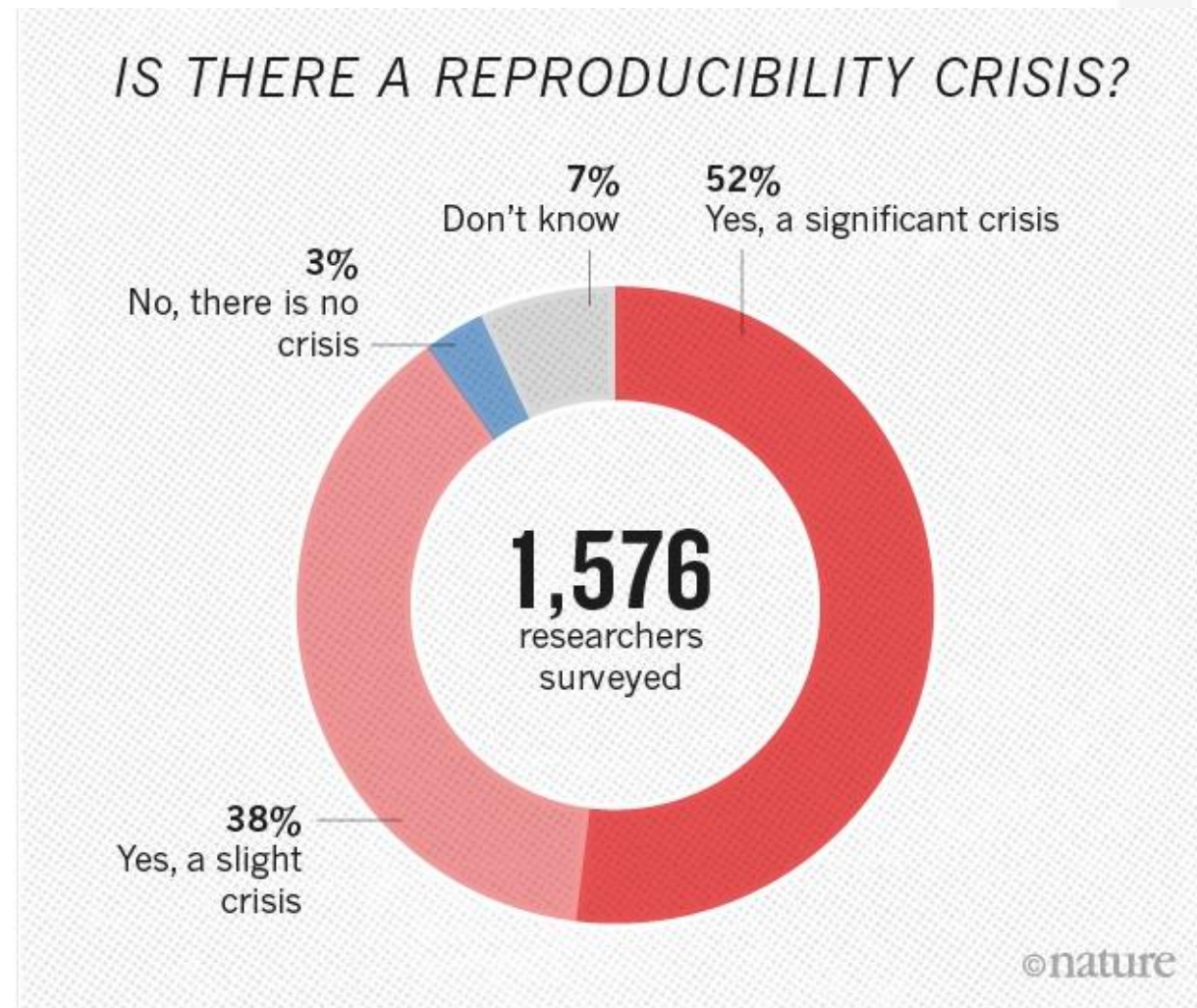
Table 1. Examples of Some Reported Reproducibility Concerns in Preclinical Studies

Author	Field	Reported Concerns
Ioannidis et al (2009) ²²	Microarray data	16/18 studies unable to be reproduced in principle from raw data
Baggerly et al (2009) ²³	Microarray data	Multiple; insufficient data/poor documentation
Sena et al (2010) ²⁴	Stroke animal studies	Overt publication bias: only 2% of the studies were negative
Prinz (2011) ¹	General biology	75% to 80% of 67 studies were not reproduced
Begley & Ellis (2012) ²	Oncology	90% of 53 studies were not reproduced
Nekrutenko & Taylor(2012) ²⁵	NGS data access	26/50 no access to primary data sets/software
Perrin (2014) ²⁶	Mouse, in-vivo	0/100 reported treatments repeated positive in studies of ALS
Tsilidis et al (2013) ²⁷	Neurological studies	Too many significant results, overt selective reporting bias

Begley, C. G. & Ioannidis, J. P. Reproducibility in science: improving the standard for basic and preclinical research. *Circ. Res.* 116, 116–126 (2015).



Scope of Problem: Science Researchers



Disciplines Represented

- Biology
- Chemistry
- Physics
- Medicine
- Environmental Sciences

Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454 (2016).



Scope of Problem: Social Sciences

- 21 high-powered replications of social science experiments
 - Failed to replicate results of more than a third of the studies
 - significantly weaker evidence for remainder compared to original studies

Testing the reproducibility of social science research

<https://penntoday.upenn.edu/news/testing-reproducibility-social-science-research>

- The Reproducibility Project: Psychology
 - Replicate effects of 100 psychology studies
 - 97% of original studies produced statistically significant results
 - only 36% of the replication studies did

Jeffrey R. Stevens. Replicability and Reproducibility in Comparative Psychology, Front. Psychol., 26 May 2017 | <https://doi.org/10.3389/fpsyg.2017.00862>



Data Reproducibility Initiatives in the Health Sciences and Social Sciences



AUGUSTA UNIVERSITY
University Libraries

Research Reproducibility Initiatives



- Pre-registering studies
- Registered Reports
 - Conduct peer review prior to data collection
 - Provisionally accepted for publication if authors follow through
 - 162 journals participate with submission option/special issue

BMC Biology

BMC Ecology

BMC Medicine

BMJ Open Science

Brain and Behavior

British Journal of Clinical Psychology

British Journal of Developmental Psychology

British Journal of Educational Psychology

British Journal of Health Psychology



<https://cos.io/rr/>



AUGUSTA UNIVERSITY
University Libraries

Research Reproducibility Initiatives

- NIH Common Data Elements (CDE) Repository
 - Defines data elements required/recommended
 - NIH Institutes and other organizations
 - Includes data collection forms



<https://cde.nlm.nih.gov/>

Browse by Classification Browse by Topic Export Search		
Analytical, Diagnostic and Therapeutic Techniques, and Equipment Diagnosis (3503) Therapeutics (36) Equipment and Supplies (5)	Anatomy Cardiovascular System (7) Respiratory System (7)	Anthropology, Education, Sociology, and Social Phenomena Social Sciences (17) Human Activities (8)
Disciplines and Occupations Natural Science Disciplines (220)	Diseases Pathological Conditions, Signs and Symptoms (73) Nervous System Diseases (54)	Health Care Health Care Facilities, Manpower, and Services (36) Environment and Public Health (9)
Humanities Humanities (9)	Information Science Information Science (8)	Phenomena and Processes Musculoskeletal and Neural Physiological Phenomena (78) Genetic Phenomena (3)



Research Reproducibility Initiatives

- Data Management Guidelines/Plans
 - NIH Data Sharing Policy (2003)
 - Data Questions
 - (1) What data will be shared?
 - (2) Who will have access to the data?
 - (3) Where will the data to be shared be located?
 - (4) When will the data be shared?
 - (5) How will researchers locate and access the data?
 - NSF Data Management Plan (2011)
 - Biological Sciences
 - Computer and Information Science
 - Education and Human Resources
 - Mathematical and Physical Sciences
 - Social, Behavioral and Economic Sciences



Research Reproducibility Initiatives

- Clinical Trials Registration and Results Information Submission (2017)

- Required reporting data
 - Participant flow
 - Baseline characteristics
 - Outcome measures
 - Adverse events

ClinicalTrials.gov

[Find Studies](#) ▼ [About Studies](#) ▼ [Sub](#)

[Home](#) > [Submit Studies](#) > [Go to PRS](#)

SUBMIT STUDIES

[Why Should I Register and Submit Results?](#)

[FDAAA 801 and the Final Rule](#)

[How to Apply for an Account](#)

[How to Register Your Study](#)

[How to Edit Your Study Record](#)

[How to Submit Your Results](#)

[Frequently Asked Questions](#)

Go to PRS

Data providers use a Web-based data entry system called the Protocol Registration and Results submit results information for a registered study.

You must have a PRS account to register study information on ClinicalTrials.gov.

If you do not have an account, see [How to Apply for an Account](#).

To proceed to PRS, enter the day of week exactly as it appears below and continue.

Monday



FAIR Principles

What is FAIR DATA?



Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

FINDABLE



Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.

ACCESSIBLE



Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.

INTEROPERABLE








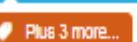


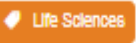






Data and collections have a clear usage licenses and provide accurate information on provenance.

REUSABLE

<https://libereurope.eu/wp-content/uploads/2017/12/LIBER-FAIR-Data.pdf>



Registry	Name	Abbreviation	Type	Subject	Domain	Taxonomy	Related Database	Related Standard	Related Policy
	ArrayExpress	ArrayExpress	Database	 Life Sciences	 Chromatin Immunoprecipitation - DNA Microarray  Chromatin Immunoprecipitation - DNA Sequencing  DNA Microarray  Gene Expression Data  Genotyping  Plus 3 more...	 All	BioSD Expression Atlas SMD DGVA FlyMine Plus 5 more...	MAGE-TAB EFO MIAME/Env MINSEQE MIAME Plus 2 more...	UCL Child Open Health Research EMBO Recommended Databases and Data Standards HRB Open Research PLOS Recommended Repositories MNI Open Research Plus 47 more...
	Gene Expression Omnibus	GEO	Database	 Life Sciences	 DNA Microarray  Gene Expression Data  Genome	 All	Nematode.net SMD ImmPort BioSample ENCODE Plus 5 more...	MIAME/Env SOFT MIAME MINiML	Springer Nature Research Data Policy Type 4 Springer Nature - Research Data Policies Springer Nature Research Data Policy Type 1 UCL Child Open Health Research EMBO Recommended

In the following recommendations:



- Resource for data and metadata standards
- Inter-related to databases and data policies

Collections and Recommendations

Contribute by adding a collection or recommendation | Any problems? Please tell us

Collections group together one or more types of resource (standard, database or policy) by domain, project or organisation. A Recommendation is a core-set of resources that are selected or endorsed data policies from journals, funders or other organizations.

Search [Search] [Reset]

48 records in view

View as Table | View as Grid

Sort by: Name

View Recommendations Only

Record Type

- Journal: 14
- Project: 14
- Organization: 12
- Domain: 3
- Society: 3

Registry	Name	Abbreviation	Type	Subject	Domain	Taxonomy	Related Database	Related Standard	Related Policy	In Collection/Recommendation	Status
	Chemistry	N/A	Domain	Chemistry	None	All	GlycoNA/I canSAR COD MINAS Polbase Plus 40 more...	nmrCV CHEBI CHEMINF SOPHARM CDISC LAB Plus 29 more...	Springer Nature Research Data Policy Type 4 ACS - ACR Policy Springer Nature Research Data Policy Type 1 ACS - CR Policy Hindawi Research Data Policy Plus 8 more...	N/A	N/A
	Computational Modeling	N/A	Domain	Biomedical Science Environmental Sciences Life Sciences	Mathematical Model	All	PMR JWS Online ProbOnto VPR	BioPAX CellML NeuroML SBGN SED-ML Plus 17 more...	None	N/A	N/A
	DNA Microarray	N/A	Domain	Biomedical Science Life Sciences	DNA Microarray Gene Expression Data	All	ArrayExpress GEO	MINIML MAGE-TAB MIAME	None	N/A	N/A



Research Reproducibility Initiatives – Science Data Journals

The logo for Scientific Data, featuring the text "SCIENTIFIC DATA" in white on a dark blue background, with a small graphic of binary code to the right.

- Data Descriptor articles
- New or published studies
- Prior published with new content
- published fully open access.
- Peer review process

The logo for the Journal of open health data, featuring the text "Journal of open health data" in white and orange on a dark background with a cityscape image.

- Health datasets with
- High reuse potential
- Archived, preserved, and openly available

- Title (110 characters maximum, including whitespaces)
- Abstract (170 words maximum, no references)
- Background & Summary (700 words maximum)
- Methods (unlimited length)
- Data Records (unlimited length)
- Technical Validation (unlimited length)
- Usage Notes (unlimited length)
- References
- Figures (generally no more than three per manuscript)
- Tables (generally no more than ten per manuscript)

Licence

PhARaoH requires new users to register and agree to the conditions of use which can be found at <http://www.lboro.ac.uk/research/mi-lab/research/pharaoh/pharaohconditionsofuse/>. Instructions for distribution to third parties are also outlined.

Accessibility criteria

Visit the PhARaoH webpage at <http://www.lboro.ac.uk/research/mi-lab/research/pharaoh/>.

Repository location

Data is available upon request from Movement Insights Lab at <http://www.lboro.ac.uk/research/mi-lab/research/pharaoh/pharaohconditionsofuse/>. Manuscripts are required to be sent to the PhARaoH study team before submission to peer review journals.

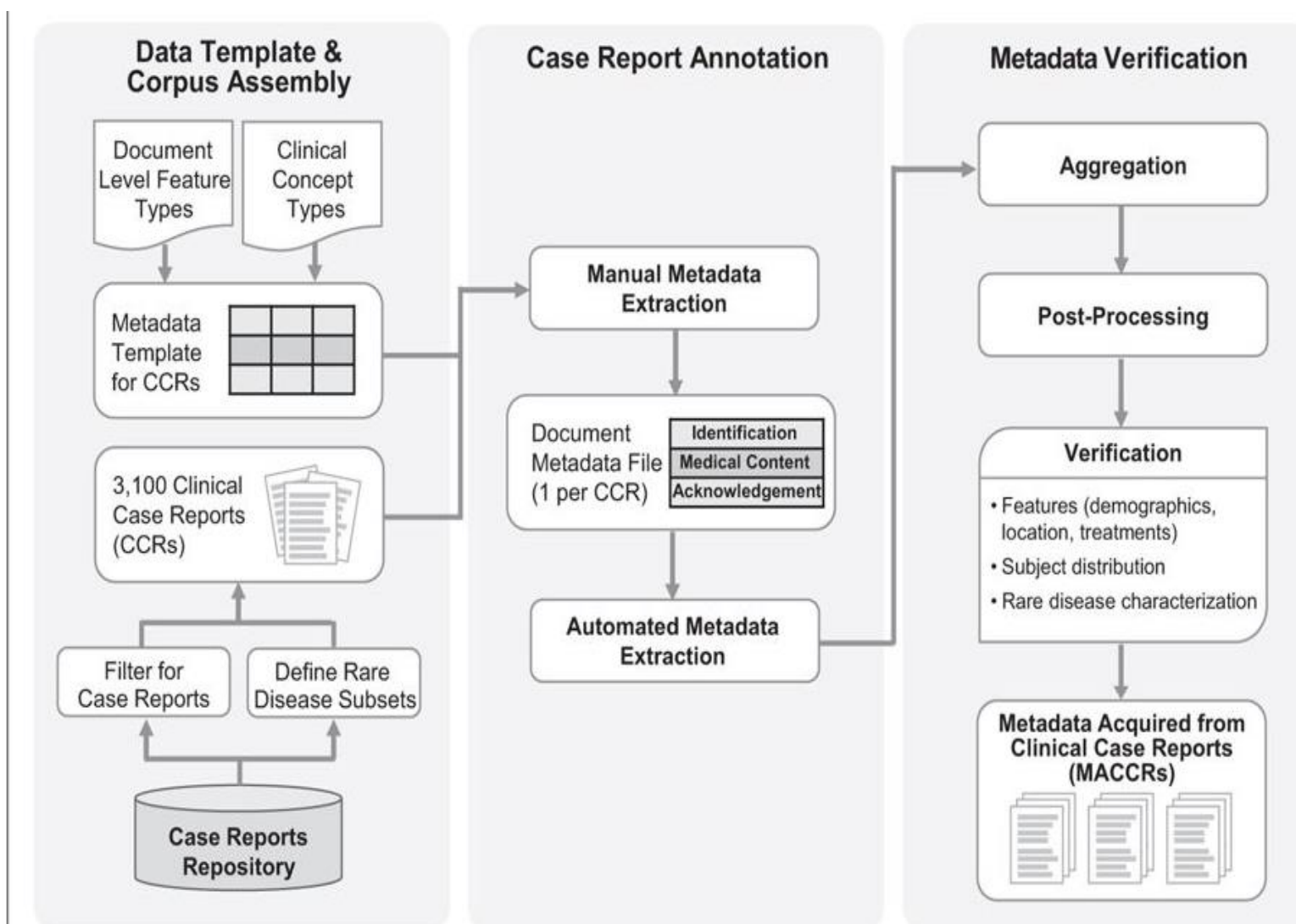


AUGUSTA UNIVERSITY
University Libraries

A reference set of curated biomedical data and metadata from clinical case reports

J. Harry Caulfield, Yijiang Zhou, Anders O. Garlid, Shaun P. Setty, David A. Liem, Quan Cao, Jessica M. Lee, Sanjana Murali, Sarah Spendlove, Wei Wang, Li Zhang, Yizhou Sun, Alex Bui, Henning Hermjakob, Karol E. Watson & Peipei Ping ✉

Scientific Data **5**, Article number: 180258 (2018) | [Download Citation](#) 



Science Data Repositories



NIH Data Repositories

- NIH-supported data repositories for data reuse
- Data from NIH-funded investigators (and others)
- Portal resources for information about biomedical data
- Single discipline, subject focused, or multidisciplinary

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
Common Fund	Epigenomics	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
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
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 relationship to genetic variation. This project will collect and analyze multiple human tissues from donors who are also densely genotyped, to assess genetic variation within their genomes. By analyzing global RNA expression within individual tissues and treating the expression levels of genes as quantitative traits, variations in gene expression that are highly correlated with genetic variation can be identified as expression quantitative trait loci, or eQTLs.	Project data only	How to access GTEx data	Yes	No	Yes	No
Common Fund	HMP DACC	The HMP DACC is a common repository for diverse human microbiome datasets and minimum reporting standards for the Common Fund Human Microbiome Project (HMP).	Project data only	How to access HMP DACC data	No	No	Yes	No
Common Fund	Illuminating Druggable Genome	The goal of the Illuminating the Druggable Genome (IDG) program is to improve our understanding of the properties and functions of proteins that are currently unannotated within the three most commonly drug-targeted protein families: G-protein coupled receptors, ion channels, and protein kinases.	Project data only	How to access IDG data	Yes	No	Yes	Yes
Common Fund	International Mouse Phenotyping Consortium (IMPC)	Phenotype data on knockout mouse lines.	How to submit data to IMPC	How to access IMPC data	Yes	No	Yes	Yes



NIH Data Repositories: Portal

- Immunology Database and Analysis Portal

- Analysis tools
- Publication links
- Resource links

EBioMedicine

Published by THE LANCET

Available online 27 December 2018

open access

In Press, Corrected Proof ?

Research paper

Development and validation of an immune gene-set based Prognostic signature in ovarian cancer

Sipeng Shen ^{a, b, c, 1}, Guanrong Wang ^{d, 1}, Ruyang Zhang ^{a, b, c}, Yang Zhao ^{a, b, c}, Hao Yu ^c, Yongyue Wei ^{a, b, c}, Feng Chen ^{a, b, c} 

 [Show more](#)

<https://doi.org/10.1016/j.ebiom.2018.12.054>

[Get rights and content](#)

Under a Creative Commons license

Abstract

Background

Ovarian cancer (OV) is the most lethal **gynecological cancer** in women. We aim to develop a generalized, individualized immune prognostic signature that can stratify and predict **overall survival** for ovarian cancer.

Methods

The gene expression profiles of ovarian cancer tumor tissue samples were collected from 17 public cohorts, including 2777 cases totally. Single sample **gene set enrichment** (ssGSEA) analysis was used for the immune genes from ImmPort database to develop an immune-based prognostic score for OV (IPSOV). The signature was trained and validated in six independent datasets ($n = 519, 409, 606, 634, 415, 194$).



IMMPORT

BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY

ImmPort is funded by the NIH, NIAID and DAIT in support of the NIH mission to share data with the public. Data shared through ImmPort has been provided by NIH-funded programs, other research organizations and individual scientists ensuring these discoveries will be the foundation of future research.



Private Data

Upload
Validator
Search



Shared Data

Data Model
Search/Download
Gene Lists



Data Analysis

Analysis Workflow
Automated Clustering
Tutorials



Resources

Tutorials
Documentation
Publications



immuneXpresso

immuneXpresso (iX) allows to learn about reported literature interactions between various cell types and regulatory molecules (cytokines chemokines).



AUGUSTA UNIVERSITY
University Libraries

NIH Data Repositories: IMMPort

- Create account
- Meets journal standards
- Data sharing & access agreement required

Data Summary

Updated: January 2019, Version: DR29

Studies	361
Subjects	54258
Protocols	1504
Experiments	1535
Total Results	1827675
ELISA Results	247278
ELISPOT Results	69709
Flow Cytometry Results	933214
Gene Expression Results	125302
HAI Results	22592
HLA Typing Results	71294
Multiplex Results	337184
Neutralizing Antibody Results	21102

SDY404 - Immunologic and genomic signatures of influenza vaccine response - 2011 (see companion studies SDY63, SDY400, SDY520)

Summary	Design	Adverse Event	Assessment	Interventions	Medications	Substance	Demographics	Lab Tests	Mechanistic Assays	Study Files
+ -										
▼ Summary										
Accession	SDY404									
DOI	10.21430/M3GWQRC8DT									
Title	Immunologic and genomic signatures of influenza vaccine response - 2011 (see companion studies SDY63, SDY400, SDY520)									
PI	David Hafler - Yale									
Type										
Condition Studied	Immunologic and genomic signatures of influenza vaccine response									
Brief Description	Project 1: Immunologic and genomic signatures of influenza vaccine response - year2 2011									
Start Date	2011-01-01									
Detailed Description	Project 1: Immunologic and genomic signatures of influenza vaccine response - year2 2011									
Objectives	Identification of immunologic, gene expression, and genetic loci signatures of vaccine response, and effects of aging and functional status on these signatures.									
Endpoints	B cells (ASC, Memory); T cells (Ag specific CD8 and CD4, memory T cell subsets, Tregs, exhausted/senescent); in vitro activation of flu-specific T cells; Monocyte/DC stim by flu or through TLRs; innate cell TLR expression; Luminex assays of cytokines post vaccine; Illumina HT12 bead chip and Nanostring.									
Gender Included	Both									
Subjects Number	72									
Download Packages	Study Download Packages									

Sponsored by:

- National Institute of Allergy and Infectious Diseases (NIAID)
- National Institutes of Health (NIH)
- Health and Human Services (HHS)

Nature Scientific Data's
Recommended Data Repository
Cytometry & Immunology

PLOS ONE
Recommended
Data Repository

[Contact Us](#) | [Privacy Policy](#) | [Disclaimer](#) | [Accessibility](#)



AUGUSTA UNIVERSITY
University Libraries

NIH Data Repositories

National Sleep Research Resource

Free research data and tools.

[NSRR](#) [About](#) [Datasets](#) [Tools](#) [Forum](#) [Blog](#)

[Search](#) [Sign in](#)

What interests you?

POLYSOMNOGRAPHY



ACTIGRAPHY



DATASETS



SHARING DATA



Not sure? View our [most popular](#) datasets.

Tools

Want to do more with the data?



EDF Visualization

[EDF Viewer](#) · [EDFbrowser](#) · [Polyman](#) · [Altamira](#)



Signal Processing

[Data Chromatix](#) · [HRV Toolkit](#) · [SpectralTrainFig](#) · [SpiSOP](#)



Learn

3.0 ★ [Count Apneas in EDF Annotations](#)
3.0 ★ [Download EDFs programmatically](#)
3.0 ★ [Read EDF Header and Signal Header](#)

Know of another great tool? [Tell us about it.](#)



About

The Sleep Heart Health Study (SHHS) is a multi-center cohort study implemented by the National Heart Lung & Blood Institute to determine the cardiovascular and other consequences of sleep-disordered breathing. It tests whether sleep-related breathing is associated with an increased risk of coronary heart disease, stroke, all cause mortality, and hypertension. In all, 6,441 men and women aged 40 years and older were enrolled between November 1, 1995 and January 31, 1998 to take part in SHHS Visit 1. During exam cycle 3 (January 2001- June 2003), a second polysomnogram (SHHS Visit 2) was obtained in 3,295 of the participants. CVD Outcomes data were monitored and adjudicated by parent cohorts between baseline and 2011. More than 130 manuscripts have been published investigating predictors and outcomes of sleep disorders. [Click here for a full description of SHHS.](#)

Read more about SHHS at [ClinicalTrials.gov \(NCT00005275\)](#).

Data overview

[/datasets](#) (introduction)

Core data from SHHS Visit 1 and Visit 2, as well as longitudinal CVD Outcomes tracking.

[/polysomnography](#) (introduction)


Overnight polysomnography (PSG) data from SHHS Visit 1 and Visit 2.

Protocols and manuals

- [SHHS1 Protocol](#)
- [SHHS1 Manual of Operations](#)
- [SHHS1 Forms](#)
- [SHHS2 Protocol](#)
- [SHHS2 Manual of Operations](#)
- [SHHS2 Forms](#)
- [SHHS CVD Outcomes Protocol](#)
- [Sleep Reading Center Manual of Operations](#)

SHHS

mop


 [03-data-analysis-tip-sheet...](#)


 [04-dataset-introduction.md](#)

 [05-polysomnography-intro...](#)

 [06-reliability-shhs1.md](#)

 [07-reliability-shhs2.md](#)

 [08-equipment-shhs1.md](#)

 [09-equipment-shhs2.md](#)

 [10-montage-and-sampling...](#)

 [11-montage-and-sampling...](#)

 [12-eeg-spectral-analysis.md](#)

 [13-hrv-analysis.md](#)

 [CHANGELOG.md](#)

 [PUBLICATIONS.md](#)

 [README.md](#)

 [full-description.md](#)



shhs > files > datasets

archive

eeg-spectral-analysis

hrv-analysis

CHANGELOG.md

KNOWNISSUES.md

shhs1-dataset-0.13.0.csv

shhs2-dataset-0.13.0.csv

shhs-cvd-events-dataset-0.13.0.csv

shhs-cvd-summary-dataset-0.13.0.csv

shhs-data-dictionary-0.13.1-domains.csv

shhs-data-dictionary-0.13.1-forms.csv

shhs-data-dictionary-0.13.1-variables.csv

shhs-interim-followup-dataset-0.13.0.csv

Study Files And Documentation

Need EDFs? The Sleep Portal Viewer can open files provided by NSRR datasets, specifically, take a look at the [SHHS File Downloads!](#)

EDF Viewer Screenshot



NIH Data Repositories

TCIA Collections

TCIA is a service which de-identifies and hosts a large archive of medical images of cancer accessible for public download. The data are organized as "Collections", typically patients related by a common disease (e.g. lung cancer), image modality (MRI, CT, etc) or research focus. DICOM is the primary file format used by TCIA for image storage. Supporting data related to the images such as patient outcomes, treatment details, genomics, pathology, and expert analyses are also provided when available.

Read more

Show 100 entries

Filter table:

Collection	Cancer Type	Location	Subjects	Image Types	Supporting Data	Access	Status	Updated
ACRIN-NSCLC-EDG-PET	Non-small Cell Lung Cancer	Lung	193	PT, CT, MR, CR, DX, SC	Clinical	Limited	Complete	2019-02-06
Brain-Tumor-Progression	Brain Cancer	Brain	20	MR	Image Analyses	Public	Complete	2019-02-05
Acrin-FLT-Breast	Breast Cancer	Breast	83	PET, CT, OT	Clinical	Public	Complete	2019-01-22



Data Repositories: Multidisciplinary

The screenshot displays the Harvard Dataverse interface. At the top, the Harvard logo and 'Dataverse' name are on the left, while navigation links (Search, About, User Guide, Support, Sign Up, Log In) are on the right. Below the header, a metrics bar shows '5,928,798 Downloads'. A search bar with the placeholder 'Search this dataverse...' is followed by 'Find' and 'Advanced Search' buttons, and an 'Add Data' button. The left sidebar contains filters for 'Dataverses (479)', 'Datasets (2,144)', and 'Files (0)'. Under 'Dataverse Category', it lists 'Research Project (208)', 'Researcher (117)', 'Organization or Institution (54)', 'Research Group (43)', and 'Journal (12)'. The 'Metadata Source' section lists 'Harvard Dataverse (2,458)' and 'Harvested (165)'. The 'Publication Year' section lists years from 2018 to 2019 with their respective counts. The 'Subject' section shows 'Medicine, Health and Life Sciences (2,623)' as the selected filter, with other subjects like 'Social Sciences (713)' and 'Physics (145)' listed below. The main content area shows '1 to 10 of 2,623 Results' for the selected subject. It lists three results: 1. 'Full Speed Ahead' on Hospital Job (Oct 18, 2018 - VA CWM History Dataverse) by Casella, Jessie, 2018. 2. 'Raw rs-fMRI files for Static Hypnosis' (Apr 20, 2016 - Static Hypnosis Dataverse) by Barrios, Fernando, 2016. 3. 'Unilateral lesion of dopamine neurons induces grooming asymmetry in the mouse' (May 30, 2015 - Assunta Pelosi Dataverse) by Pelosi, Assunta, 2015. Each result includes a document icon, a title, a date, a source, and a brief description. A fourth result is partially visible at the bottom.

HARVARD
Dataverse

Search ▾ About User Guide Support Sign Up Log In

Metrics 5,928,798 Downloads

Contact Share

Search this dataverse... Find Advanced Search + Add Data

☒ **Dataverses (479)**
☒ **Datasets (2,144)**
☐ **Files (0)**

Dataverse Category
Research Project (208)
Researcher (117)
Organization or Institution (54)
Research Group (43)
Journal (12)
More...

Metadata Source
Harvard Dataverse (2,458)
Harvested (165)

Publication Year
2018 (793)
2016 (680)
2017 (540)
2015 (269)
2019 (106)
More...

Subject
Medicine, Health and Life Sciences (2,623)
✕
Social Sciences (713)
Physics (145)
Computer and Information Science (134)
Earth and Environmental Sciences (117)
More...

Subject: Medicine, Health and Life Sciences ✕

1 to 10 of 2,623 Results **Sort ▾**

"Full Speed Ahead" on Hospital Job
Oct 18, 2018 - VA CWM History Dataverse
Casella, Jessie, 2018, "Full Speed Ahead" on Hospital Job", <https://doi.org/10.7910/DVN/NHTZXQ>, Harvard Dataverse, V1
Daily Hampshire Gazette, July 28, 1922. Head of construction firm explains what will be done to clear site and start building foundations.

"Raw rs-fMRI files for Static Hypnosis"
Apr 20, 2016 - Static Hypnosis Dataverse
Barrios, Fernando, 2016, "Raw rs-fMRI files for Static Hypnosis", <https://doi.org/10.7910/DVN/XE8NLW>, Harvard Dataverse, V1
Raw rs-fMRI files for Static Hypnosis experiment "Brain functional connectivity of hypnosis without targeted suggestion. An intrinsic hypnosis rs-fMRI study" Pablo Vazquez, Susan Whitfield-Gabrieli, Clemens Bauer, Fernando A Barrios, (2017) submitted to Neuropsychologia, NSY-D-17...

"Unilateral lesion of dopamine neurons induces grooming asymmetry in the mouse". A. Pelosi, J.A. Girault, D. Hervé (2015)
May 30, 2015 - Assunta Pelosi Dataverse
Pelosi, Assunta, 2015, "Unilateral lesion of dopamine neurons induces grooming asymmetry in the mouse". A. Pelosi, J.A. Girault, D. Hervé (2015)", <https://doi.org/10.7910/DVN/1UK7K1>, Harvard Dataverse, V1, UNF:6:7E7TPjfkVivyujNmI9ZCcw== [fileUNF]
Examples of video used to analyze grooming behaviour in mice unilaterally lesioned by 6-OHDA in the medial forebrain bundle. Animals were video recorded after the lesion (in the first cohort of animals), before surgery, after surgery and after surgery and L-DOPA acute treatment (...)



(Partial) Replication Data for: An empirical evaluation of the causal impact of NAAQS nonattainment designations on particulate pollution and health
Jan 3, 2017 - Impact of National Ambient Air Quality Standards nonattainment designations on particulate pollution and health
Zigler, Cory, 2017, "(Partial) Replication Data for: An empirical evaluation of the causal impact of NAAQS nonattainment designations on particulate pollution and health", <https://doi.org/10.7910/DVN/ZAYLFA>, Harvard Dataverse, V1, UNF:6:w5afM9qjNg6BGRISwQ+LNw== [fileUNF]
This is the analysis data file used for the paper, complete with propensity score estimates but with all Medicare variables set to zero. This file can be used to reconstruct basic data summaries reported in the paper.

<https://dataverse.harvard.edu/>



 Metrics

13 Downloads

 Contact  Share



"Raw rs-fMRI files for Static Hypnosis" Version 1.1

Barrios, Fernando, 2016, "'Raw rs-fMRI files for Static Hypnosis'", <https://doi.org/10.7910/DVN/XE8NLW>, Harvard Dataverse, V1

 Cite Dataset ▼

 Learn about Data Citation Standards.

Description

Raw rs-fMRI files for Static Hypnosis experiment "Brain functional connectivity of hypnosis without targeted suggestion. An intrinsic hypnosis rs-fMRI study" Pablo Vazquez, Susan Whitfield-Gabrieli, Clemens Bauer, Fernando A Barrios, (2017) submitted to Neuropsychologia, NSY-D-17-00846R2. Minimum data set, 5 females and 5 males highly hypnotizable subjects according to the Stanford test. Raw resting state fMRI data in two conditions control alert state and hypnotic state regular ayes closed. (2016-04-20)

Subject

Medicine, Health and Life Sciences

Keyword

rs-fMRI

Notes

Data base in BIDS structure, file list in the filelist file, all structural images have been "defaced" with pydeface (<https://github.com/poldracklab/pydeface> data acquisition and experimental conditions are in the Neuropsychologia article under revision. Complete dataset will be deposited at the INDI database http://fcon_1000.projects.nitrc.org/indi/pro/unam_barrios_hypnosis_index.html

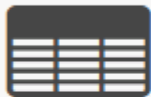
Files

Metadata

Terms

Versions

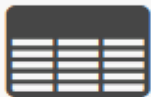




DATASET

Clinical features of papillary or follicular thyroid cancer with bone ...

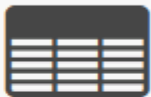
Jen-Der Lin ▾ 09/03/2017



DATASET

The distribution of bone metastases in patients with bladder cancer.

Ping Huang ▾ 13/09/2017



DATASET

The risk factors for detecting bone metastases in patients with bladder cancer.

Ping Huang ▾ 13/09/2017

	A	B	C
1	Clinical Characteristic	All Patients	Cancer Mortality vs. Non-cancer Mortality
2			Cancer Mortality
3	Patient number	131 (100.0)	88 (67.2)
4	Group A	90 (68.7)	69 (76.7%)
5	Group B	41 (31.3)	19 (46.3%)
6	Gender, Female	94 (71.8)	68 (77.3)
7	Age at diagnosis (year)	57.4 ± 15.0	60.5 ± 13.1
8	Mean tumor size (cm)	4.1 ± 2.	4.7 ± 3.3
9	Median post-operative serum Tg level after 1 month (ng/mL)	300	472
10	(range)	(0.0–141970.0)	(1.4–141970.0)
11	Multifocality	15 (11.5)	9 (10.2)
12	Operative method		
13	Total thyroidectomy	87 (66.4)	56 (63.6)
14	Less than total thyroidectomy	44 (33.6)	32 (36.4)



Open Data Repository

- figshare
 - Data sets and meta data
 - Scholarly presentations/articles
 - Limit to dataset in results

FIGSHARE PROVED TO BE A GOOD IDEA. SINCE 2012 WE HAVE OFFERED...

26 million +
page views

7.5 million +
downloads

800,000 +
user uploads

2 million +
articles

500,000 +
collections

5,000 +
projects

<https://figshare.com>



Dryad: Journal Publication Data



Open data best practices:

How to make your Dryad data package as reusable as possible

Submit data now

[How and why?](#)

Search for data

Enter keyword, author, title, DOI, etc

Go

[Advanced search](#)

Browse for data

Recently published

Popular

Most Downloaded Items

Number of Downloads

Hansen B, Jespersen SN (2016) Data from: Data for evaluation of fast kurtosis imaging, b-value optimization and exploration of diffusion MRI contrast. *Scientific Data*
<https://doi.org/10.5061/dryad.9bc43>

38653

Latest from @datadryad

Tweets by @datadryad

Dryad Retweeted



Eco Monographs
@ESAMonographs

#New in @ESAMonographs with
#OpenData from @datadryad

The study of #reservoir delta-backwaters is
a new frontier in river
ecology#ecosystemservice #restoration



AUGUSTA UNIVERSITY
University Libraries

Look up your journal

Show all amounts in USD\$ ▼

Dryad welcomes data submissions associated with any published article. We work with journals to [integrate article submission with data submission](#), thus streamlining the process, as well as encourage partners to sponsor their authors' [Data Publishing Charges](#) (DPCs). All journals with integrated data submission and/or sponsored DPCs are listed below.

The Data Publishing Charge is \$120. If you see an organization listed under "Payment" below, the DPC is covered. In addition, our [institutional sponsors](#) fund submissions by their affiliated researchers.

For large data packages (over 20GB), submitters will be charged \$50 for each additional 10GB, or part thereof, beyond 20GB. (Packages between 20 and 30GB = \$50, between 30 and 40GB = \$100, and so on).

Data from: Increased soluble IL-7 receptor concentrations associate with improved IL-7 therapy outcomes in SIV-infected ART-treated Rhesus macaques

Steele AK, Carrasco-Medina L, Sodora DL, Crawley AM

Date Published: December 22, 2017

DOI: <https://doi.org/10.5061/dryad.72c40>



Files in this package

Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the [Dryad Terms of Service](#). To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.

Title	Raw data for online upload
Downloaded	20 times
Description	This file contains the raw data for all of the data figures in this manuscript.
Download	Raw data for online upload.xlsx (58.53 Kb)
Details	View File Details



Data Reproducibility Components and Challenges in the Arts and Humanities



Scope of Problem: Arts and Humanities

- What does data mean?
 - My work is not data (Thoegersen, 2018; Buys and Shaw, 2015)
 - Object itself
 - Details collected from the object
- Is data reproducibility even possible? (Peels, 2019)



Data Reproducibility Components

Health Sciences, Social Sciences, Arts and Humanities

- Sharing analytical data sets (accessibility)
- Relevant metadata
- Analytical code
- Related software
- Verify data is accurately represented

Also:

- Attribution
 - Copyright/Licensing
 - Citation



Data Copyright

- Data is considered *discoverable fact* and therefore **not under copyright** in the U.S. (<https://www.copyright.gov/docs/regstat092303.html>)
 - Compilations are under copyright, but must meet certain requirements
 - patents
 - trademarks
 - trade secrets
- So, how can I protect my data? → How do I **get credit** for my data?
 - Obtain a license to grant permissions on its reusability
 - Creative Commons Zero (CC0)



Data Citation Principles

1. Importance

Data should be considered legitimate, citable products of research. Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications[1].

2. Credit and Attribution

Data citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the data, recognizing that a single style or mechanism of attribution may not be applicable to all data[2].

3. Evidence

In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited[3].

4. Unique Identification

A data citation should include a persistent method for identification that is machine actionable, globally unique, and widely used by a community[4].

5. Access

Data citations should facilitate access to the data themselves and to such associated metadata, documentation, code, and other materials, as are necessary for both humans and machines to make informed use of the referenced data[5].

6. Persistence

Unique identifiers, and metadata describing the data, and its disposition, should persist -- even beyond the lifespan of the data they describe[6].

7. Specificity and Verifiability

Data citations should facilitate identification of, access to, and verification of the specific data that support a claim. Citations or citation metadata should include information about provenance and fixity sufficient to facilitate verifying that the specific timeslice, version and/or granular portion of data retrieved subsequently is the same as was originally cited[7].

8. Interoperability and Flexibility

Data citation methods should be sufficiently flexible to accommodate the variant practices among communities, but should not differ so much that they compromise interoperability of data citation practices across communities[8].

<https://www.force11.org/datacitationprinciples>



Data Citation

Principle 2
- **Credit and Attribution:**
Such as authors, repositories or other distributors and contributors.

← **Author(s)**, Year, Dataset Title, **Global Persistent Identifier**, →
← **Data Repository or Archive**,
Version

Principles 4, 5, 6
- **Unique Identification, Access, Persistence:**
A unique, persistent identifier, such as a DOI or Handle, that provides access to metadata.

↓

Principle 7 - Specificity and verification:
Such as the specific version used. Versioning or timeslice information should be supplied with any updated or dynamic dataset.



Data Citation

Painters network

Michael Kitromilidis, Tim Evans & Tim S. Evans

Fileset published 2019 via Figshare

Data associated with paper: "Community detection with metadata in a network of influence between painters"



<https://doi.org/10.6084/m9.figshare.5419216> Cite

Painters network

APA

Harvard

MLA

Vancouver

Chicago

IEEE

BibTeX

RIS

Kitromilidis, M., Evans, T., & Evans, T. S. (2019). Painters network. Figshare.

<https://doi.org/10.6084/M9.FIGSHARE.5419216>

Copy to Clipboard



Data Reproducibility Initiatives in the Arts and Humanities



AUGUSTA UNIVERSITY
University Libraries

Research Reproducibility Initiatives

- Open Government Plans
 - Publishing grant award information online for public access
 - Requiring data management plans for some grant applications
 - Requiring the research that derives from grant funding be made openly accessible



Research Reproducibility Initiatives

- Data Management Plans
 - Types of data produced
 - Data preservation and storage
 - Mechanism for sharing data
 - Timeline for public data access
- NEH white papers post research
 - Data management challenges



Research Reproducibility Initiatives – Humanities Data Journals



- **Metapapers** - datasets with high reuse potential
- **Full length research papers** - aspects or methods used to create, process, evaluate, or curate humanities data.
- Peer-reviewed, open access

Open Access Policy

Authors of articles published remain the copyright holders and grant third parties the right to use, reproduce, and share the article according to the **Creative Commons** license agreement.

Authors are encouraged to publish their data in recommended repositories. For a list of generic and subject specific repositories that meet our peer review criteria, see [here](#).

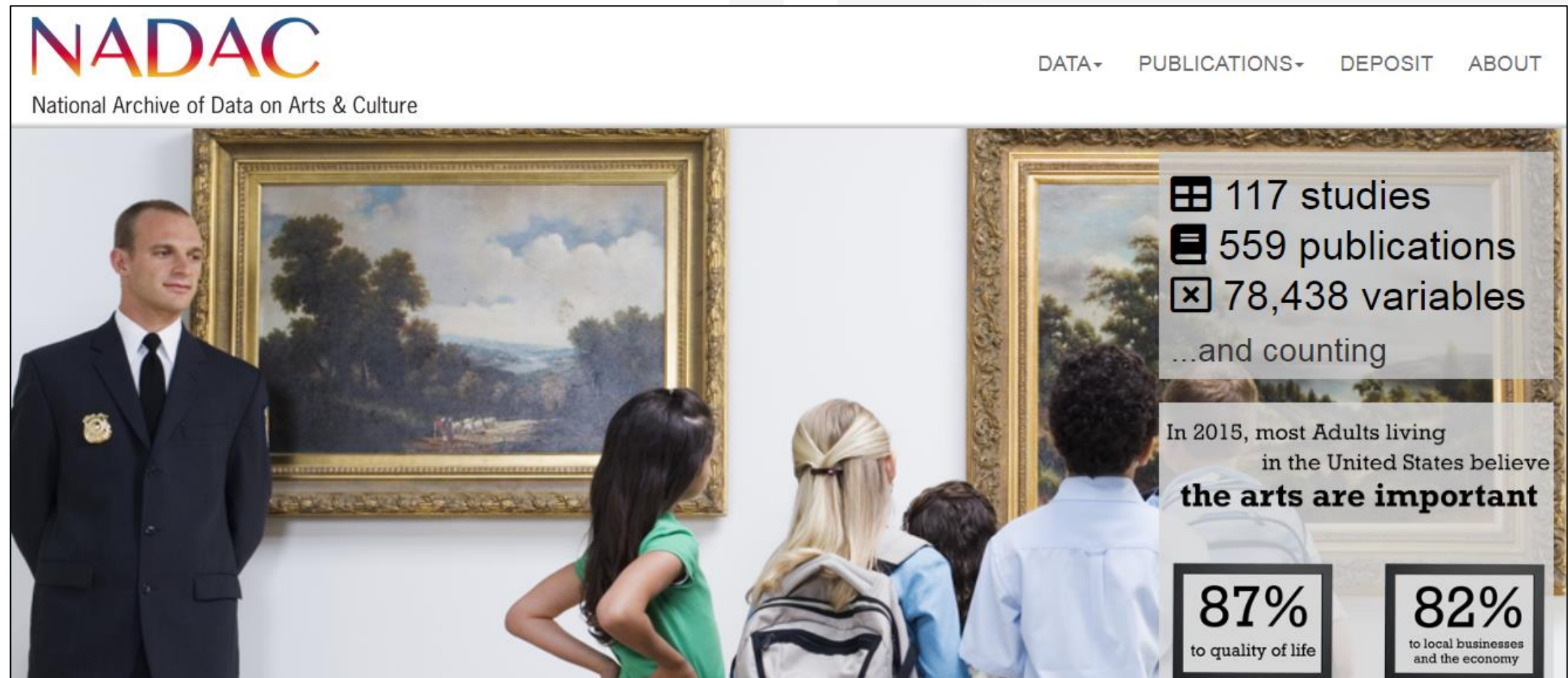


Repositories for the Arts and Humanities



AUGUSTA UNIVERSITY
University Libraries

Data repositories for the arts and humanities



<https://www.icpsr.umich.edu/icpsrweb/NADAC/index.jsp>

- Funded by NEA
 - openly accessible data
- Included in the Inter-university Consortium for Political and Social Research (ICPSR)



Data repositories for the arts and humanities



National Archive of Data on Arts & Culture

Filters

SUBJECT TERMS

arts (51)
arts participation (45)
music (45)
income (36)
arts education (35)

[view all](#)

GEOGRAPHY

DATA FORMAT

TYPE OF ANALYSIS

TIME PERIOD

RECENT RELEASES

SEARCH RESULTS

117 results.

[search tips](#) ▼

Studies (117)

Data-related Publications (560)

Variables (78,438)

Sort by: Released/Updated ▼

⏮ ⏪ 1 2 3 ⏩ ⏭

Study Title/Investigator

Released/Updated

1. [Survey of Public Participation in the Arts \(SPPA\), United States, 2017](#) (ICPSR 37138)
National Endowment for the Arts; United States. Bureau of the Census

2019-02-04

[more info](#)

2. [Early Childhood Longitudinal Program \(ECLS\), Including Arts and Cultural Variables](#) (ICPSR 37235)
United States Department of Education. National Center for Education Statistics

2019-01-08

[more info](#)

3. [Panel Study of Income Dynamics \(PSID\), Including Arts and Cultural Variables](#) (ICPSR 37234)
University of Michigan. Survey Research Center

2019-01-08

[more info](#)

- Data on the arts and its value and impact to the individual and community
- Data from projects funded by federal agencies and other organizations



AUGUSTA UNIVERSITY
University Libraries

Data repositories for the arts and humanities



Cultural Policy and the Arts National Data Archive

<http://www.cpanda.org/cpanda/>

- Data now included in NADAC but literature database remains active on cpanda site
- Data on the arts and cultural policy of the United States
- Funded by NEA
 - openly accessible data



AUGUSTA UNIVERSITY
University Libraries

Data repositories for the arts and humanities



Enter Search Term: within and

Optional Second Term: within and

Optional Third Term: within

Limit by:

Document Type: Date: to ex: 1999

Title: Public Participation in Classical Ballet: A Special Analysis of the Ballet Data Collected in the 1982 and 1985 Survey of Public Participation in the Arts

Keegan, Carol

Year: 1987

Document Type: book

Title: The finances of the performing arts, Volume 2: a survey of the characteristics and attitudes of audiences for theater, opera, symphony, and ballet in 12 U.S. cities

Ford Foundation

Year: 1974

Document Type: book

Title: Recruitment patterns: their impact on ballet and modern dance

Sussman, Leila

Year: 1990

Document Type: article

Title: The Stories of Serenade: Nonprofit History and George Balanchine's "First Ballet in America" (CACPS Working Paper No. 46)

Steichen, James

Year: 2012

Document Type: working paper/report

- Bibliographic database of literature pertaining to cultural policy
 - strong emphasis on empirical data collected on culture and the arts



Data repositories for the arts and humanities

Inter-university Consortium for Political and Social Research (ICPSR)



<https://www.icpsr.umich.edu/icpsrweb/>

- Multidisciplinary in scope
- Created and maintained by the Institute for Social Research at University of Michigan
- AU is a member of ICPSR
- Thematic data collections
- Replication data sets



Data repositories for the arts and humanities



Find Data

GO[view all](#)
[search tips](#) ▼

Browse

[Topics](#) / [Series](#) / [Thematic data collections](#) / [Global data](#)


- [List studies for which online analysis is available](#)
- [List replication datasets \(more info\)](#)
- [List studies with learning guides](#)


New/Updated Data Releases:


- [In the last week](#)
- [In the last month](#)
- [In the last quarter](#)
- [In the last year](#)

[Browse by Keyword](#)


Statistics

 [10,918 studies](#)

 [5,276,431 variables](#)

 [79,668 publications](#)

Most Popular Search Terms

A word cloud of popular search terms. The most prominent words are "general social survey", "add health", "MIDUS", "education", "mental health", "immigration", "income", "happiness", "crime", "health", "suicide", "gender", "race", "alcohol", "police", "diabetes", "China", "prison", "obesity", "religion", "marijuana", "terrorism", "bullying", "recidivism", "domestic violence", "monitoring the future", "pathways to desistance", "school crime supplement", "american community survey", "longitudinal", "social media", "age", "drinking", "sex", "ag", "nerd", "obesity", "prison", "religion", "marijuana", "terrorism", "bullying", "recidivism", "domestic violence", "monitoring the future", "pathways to desistance", "school crime supplement", "american community survey".



Data repositories for the arts and humanities



Filters

Subject Terms ▼

Restriction Type ▼

Data Format ▼

Collection Method ▼

Data Type ▼

Time Method ▼

Time Period ▼

Recent Releases ▼

Funding Agency ▼

Thematic Collection ▼

Data Availability ▼

Investigator Affiliation ▼

Search Results

587 results.

candy

GO

VIEW ALL

[search tips](#) ▼

Studies (587)

Variables (1,341)

Series (69)

Data-related Publications (6)

ICPSR Website (0)

Sort by: Study Relevance ▼

1 2 3 4 5

Study Title/Investigator

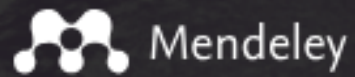
Released/Updated

1. [National Center for Early Development and Learning Multistate Study of Pre-Kindergarten, 2001-2003 \(ICPSR 4283\)](#) 2017-07-17
Clifford, Richard M.; Bryant, Donna; Burchinal, Margaret; Barbarin, Oscar; Early, Diane; Howes, Carollee; Pianta, Robert; Winton, Pam
[more info](#)
2. [New York Times/Cornell University/NY1 News New York State Poll, May 2009 \(ICPSR 26949\)](#) 2010-04-27
The New York Times; Cornell University; NY1 News
[more info](#)
3. [Evaluation of the Texas State Public School Nutrition Policy Change on Student Food Selection and Sales, School Years 2003-2004 and 2004-2005 \(ICPSR 20966\)](#) 2007-11-19
Cullen, Karen W.
[more info](#)



AUGUSTA UNIVERSITY
University Libraries

Data repositories for the arts and humanities



<https://data.mendeley.com/>

- Multidisciplinary in scope
- Elsevier product
- Open, free-to-use
 - after creating an account
- Indexes other data repositories
- Provides published data sets with DOIs

Open research data repositories in our index

4TU



ICPSR

zenodo



AUGUSTA UNIVERSITY
University Libraries

Data repositories for the arts and humanities



Filter Results

FILES

- ☐ Tabular Data (6180423)
- ☐ Image (5201877)
- ☐ Document (876507)
- ☐ Text (207976)
- ☐ File Set (165101)
- ☐ Software/Code (127410)
- ☐ Video (14374)
- ☐ Unknown File Type (13184)
- ☐ Slides (10685)
- ☐ Geospatial Data (4192)
- ☐ Sequencing Data (2448)
- ☐ Audio (1308)

REPOSITORY TYPES

- ☐ Article Repositories (5419475)
- ☐ Data Repositories (4923662)

SOURCES

- ☐ ScienceDirect (4863624)
- ☐ USGS Mineral Res. (995208)
- ☐ Zenodo (897457)
- ☐ The Cambridge Crystallographic Data Centre (CCDC) (753462)
- ☐ arXiv (555851)
- ☐ GEOROC (478309)
- ☐ PANGAEA (379832)
- ☐ ClinVar (376028)
- ☐ Apollo Cambridge (219277)
- ☐ RCSB-PDB (138270)

Somatic expressions of grief and psychosomatic illness in the works of William **Shakespeare** and his coevals

Contributors: Kenneth W. Heaton

Date: 2012-10-01

Shakespeare, famed for his insights into human nature, is exceptional ...**Shakespeare's** works. The p values refer to the significance of the chi...**Shakespeare** and those of his contemporaries, in descending order of their...**Shakespeare's** works and in those of his contemporaries. A p value of

Files:

Tabular Data (4)

Shakespeare and other English Renaissance authors as characterized by Information Theory complexity quantifiers

Contributors: Osvaldo A. Rosso, Hugh Craig, Pablo Moscato

Date: 2009-03-15

Shakespeare collected editions but whose authorship is currently in dispute...**Shakespeare**) versus word totals of the corresponding work. Both words ...**Shakespeare**; in addition we also included A Lover's Complaint, a poem ...**Shakespeare** texts two groups are clearly distinguished, the first one

Files:

Image (6) Tabular Data (5)

- Tabular, sequencing, and geospatial data



Data repositories for the arts and humanities



<https://www.jstor.org/dfr/>

- Multidisciplinary in scope
 - humanities focus
- JSTOR product
 - digital library for scholarly articles, books, and primary sources
- Datasets of JSTOR content for text mining
- Data for Research service is free for anyone
 - after creating a MyJSTOR account



Data repositories for the arts and humanities



Refine Results

Update Results

Content Type:

- ☐ Journals (12,217,575)
- ☐ Book Chapters (872,314)
- ☐ Research Reports (81,620)
- ☐ Pamphlets (25,976)

Publication Date:

From To

(yyyy or yyyy/mm or yyyy/mm/dd)

Subject:

- ☐ African American Studies (51,328)
- ☐ African Studies (148,153)
- ☐ American Indian Studies (13,428)
- ☐ American Studies (441,177)
- ☐ Anthropology (321,984)
- ☐ Aquatic Sciences (124,284)
- ☐ Archaeology (306,114)
- ☐ Architecture & Architectural History (72,864)

Data for Research

A TEXT MINING SERVICE FROM JSTOR

[DfR Home](#) | [DfR Help](#)

13,197,485 Search Results



Request Dataset

Refine Results

Update Results

Content Type:

- ☐ Journals (12,217,575)
- ☐ Book Chapters (872,314)
- ☐ Research Reports (81,620)
- ☐ Pamphlets (25,976)

Showing 1-25 of
13,197,485

Sort by: Relevance ▼

1 of 527,900 [Next](#) >

JOURNAL ARTICLE

MATERNIDAD Y TRANSGRESIONES PENALES EN EL DISTRITO FEDERAL, 1940-1950

[Martha Santillán Esqueda](#)

Historia Mexicana, Vol. 68, No. 3 (271) (ENERO-MARZO 2019), pp. 1121-1164

Download PDF



AUGUSTA UNIVERSITY
University Libraries

Data repositories for the arts and humanities



About the data

JSTOR Data for Research Search URL *

To define the parameters of your dataset, please construct a search on <https://www.jstor.org/dfr/results>. Supplying us with a DfR search URL will help to expedite delivery of your dataset. Please contact us if you encounter any problems constructing a search URL or see help here: <https://www.jstor.org/dfr/about/creating-datasets>

Per document N-grams *

- ☐ 1 gram
☐ bigrams
☐ trigrams

OCR Full text**

- ☐ Yes

***Depending on the journal and issue, this may be either text extracted from page scans via OCR or the full text as submitted digitally by publishers.*

Please describe the scope and nature of the Data to be provided including specific journals, authors disciplines, and/or keywords you would like to utilize in which to focus your query.

- Datasets of JSTOR content for text mining – request datasets

Data parameters

Limit to these publication dates

Limit to these titles (separate with commas)

Limit in another way (please describe)

Access to data

List key researchers and their affiliations

Others, if any, who are expected to have access to the Data (for example, graduate students or undergraduates)? Will other outside third parties require access to the Data? Please list their names and email addresses

Use of the data

Research

I agree to destroy the Data pursuant to the terms of the Agreement *

- ☐ Agree

Research results

Results of the research using the Data will be used as follows

Publications

If publication is intended, do you plan to specify the names of publications/publishers or institutions that are the source of the Data?

Yes

Please explain



Data repositories for the arts and humanities



<https://analytics.hathitrust.org/>

- Multidisciplinary in scope
 - humanities focus
- Partnership with HathiTrust Digital Library
 - organization of academic & research institutions who offer digitized collections
 - AU is a member
- Datasets of HathiTrust digital library content for computational text analysis
 - use datasets to run algorithms, visualization
- Use of data requires affiliation with an academic institution



Data repositories for the arts and humanities



Data Stats

# of volumes	15,722,079
# of pages	5,787,519,444
# of tokens	2,449,739,213,773
# of IC volumes	9,914,509
# of IC pages	3,005,448,348
# of IC tokens	1,777,793,828,310
# of PD volumes	5,807,570
# pd pages	2,602,212,586
# pd tokens	1,197,838,539,662

Data Worksets: user-created collections of HathiTrust volumes to be treated as data and analyzed using HTRC tools and services. Worksets are curated by researchers, and they can be shared and cited to improve reproducibility.

HTRC Algorithms: web-based, click-and-run tools to perform computational text analysis on volumes in the HathiTrust Digital Library.

Tag Cloud Creator: Create a tag cloud visualization of the most frequently occurring words in a workset, where the size of the word is displayed in proportion to the number of times it occurred.

InPhO Topic Model Explorer: export data containing the word-topic and topic-document distributions, along with an interactive visualization.



Choosing a Repository and University Libraries Resources



Choosing A Data Repository

- Required/recommended repositories
 - Funding agencies
 - Journal publication
- Repository standards
 - Metadata
 - Preservation
 - Access to data
- Licensing
 - Attribution
 - Level of sharing
 - Copyright



Finding Repositories

- Browse by subject, content type , country



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Search...

Cultural Policy and the Arts National Data Archive

CPANDA

Subject(s)

Humanities Fine Arts, Music, Theatre and Media Studies Humanities and Social Sciences

Content type(s)

Scientific and statistical data formats Structured text

Country

United States

CPANDA, the Cultural Policy and the Arts National Data Archive, is the world's first interactive digital archive of policy-relevant data on the arts and cultural policy in the United States.

Filter

[Reset all](#)

Subjects ☐
Content Types ☐
Countries ☐
AID systems ☐
Data access ☐
Data access restrictions ☐
Database access ☐
Database licenses ☐
Data licenses ☐
Data upload ☐
Data upload restrictions ☐
Enhanced publication ☐
Institution responsibility type ☐
Institution type ☐
Keywords ☐
Metadata standards ☐
PID systems ☐
Provider types ☐
Quality management ☐
Repository languages ☐
Software ☐
Syndications ☐
Repository types ☐
Versioning ☐



Finding Repositories

OpenDOAR

[Browse](#)

[Search](#)

[Statistics](#)

[Policy Tool](#)

[Our APIs](#)

[Suggest](#)

[Admin](#)

Directory of Open Access Repositories

OpenDOAR is a global directory of Open Access repositories and their policies.

Browse by Country and Region

[Countries by Region \(3890\)](#)

[Americas \(1048\)](#)

[Northern America \(634\)](#)

[United States of America \(547\)](#)

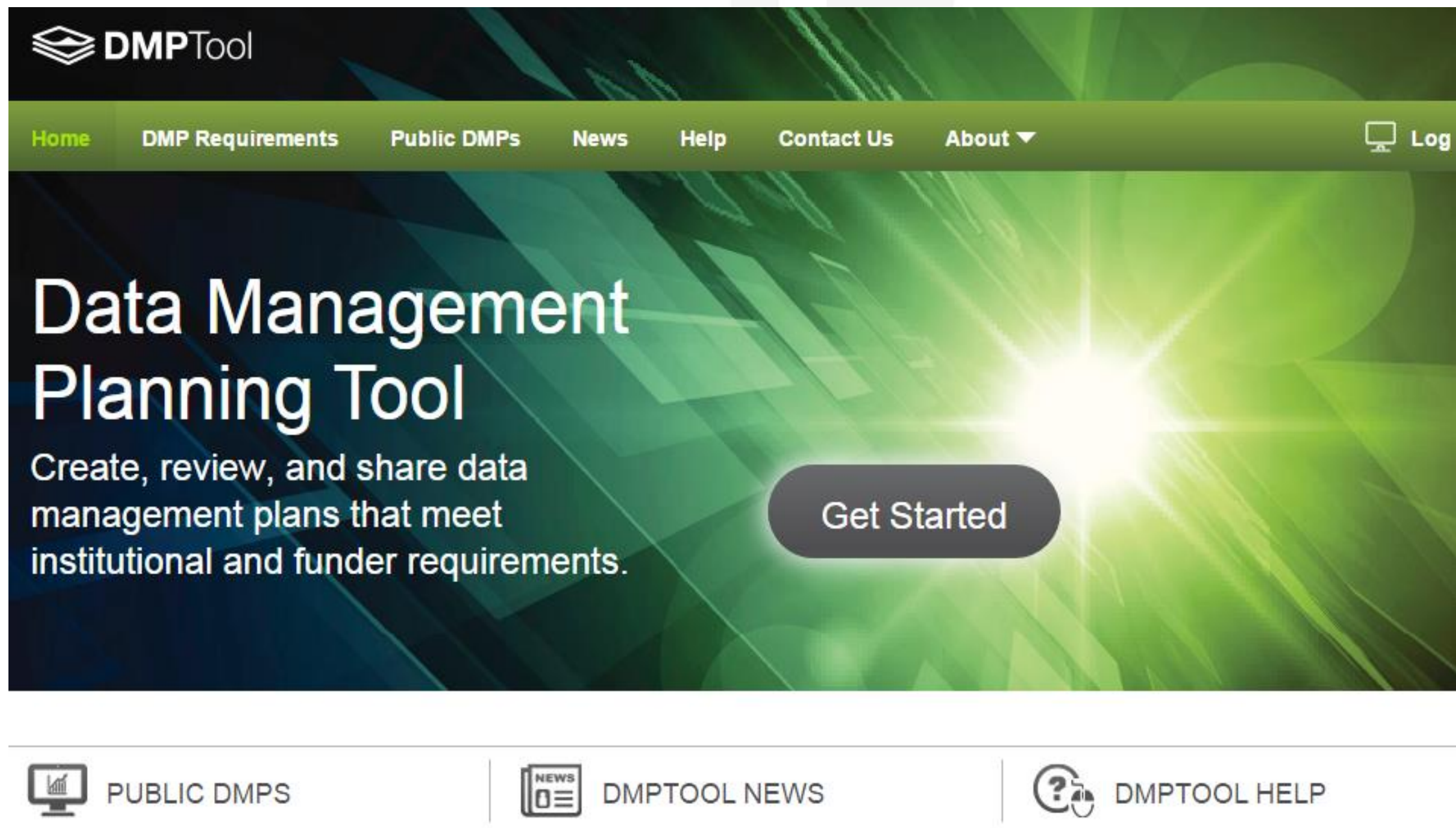
View: **by Name** | [by Software](#) | [Statistics](#)

Jump to: [1](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#)
[R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [Y](#)

Number of items at this level: 547.



Research Data Management Planning



<https://dmptool.org/>

- Augusta University participating institution
- Templates for institution and funder requirements

My Dashboard [Create plan](#)

✓ Notice: Successfully authenticated from your institutional credentials account.

My Dashboard

[Create plan](#)

The table below lists the plans that you have created, and that have been shared with you by others. You can edit, share, download, make a copy, or remove these plans at any time.

Project Title ↕	Template ↕	Edited ▼	Role	Test	Visibility	Shared	
test	NIH-GEN: Generic	10-12-2018	Owner	<input type="checkbox"/>	Private	No	Actions▼
test	NIH-GDS: Genomic Data Sharing	10-12-2018	Owner	<input type="checkbox"/>	Private	No	Actions▼

Project Details **Plan overview** Write Plan Share Download

expand all | collapse all 0/6 answered

- + Data type: human genomic data and non-human genomic data (0 / 1)
- + Data repositories (0 / 1)
- + Data submission expectations and timeline (0 / 1)
- + Informed consent and institutional certification (0 / 1)
- + Exceptions to data submission expectations (0 / 1)
- + Intellectual Property (0 / 1)

Project Details **Plan overview** Write Plan Share Download

NIH-GDS: Genomic Data Sharing

This plan is based on the "NIH-GDS: Genomic Data Sharing" template provided by National

Instructions

Data type: human genomic data and non-human genomic data

- The GDS Policy applies to all NIH-funded research that generates large-scale human research. Large-scale data include genome-wide association studies (GWAS), single metagenomic, epigenomic, and gene expression data, irrespective of funding level and support. NIH Institute or Centers (IC) may expect submission of data from smaller scale of the IC funding the research, and the utility of the data for the research community.



Data Management Planning Guide



The screenshot shows the top section of the Augusta University Libraries website. At the top is a dark blue header with the Augusta University logo (a shield with a stylized 'A' and a book) on the left, and the text 'AUGUSTA UNIVERSITY' and 'University Libraries' in white. Below the header is a breadcrumb trail: 'University Libraries / Research Guides / Data Management Planning / Home'. The main heading is 'Data Management Planning: Home'. Below this is a subheading: 'Learn how to write a successful data management plan according to funding agency requirements.' A horizontal navigation bar contains five buttons: 'Home' (highlighted in light blue), 'Funding Agency Requirements', 'How to Write a DMP', 'Resources', and 'Research Data Management Symposium'. Below the navigation bar are three large, light blue icons with corresponding text: a list icon for 'Funding Agency Requirements', a pencil icon for 'Writing a DMP', and a book icon for 'Resources'.

AUGUSTA UNIVERSITY
University Libraries

University Libraries / Research Guides / Data Management Planning / Home

Data Management Planning: Home

Learn how to write a successful data management plan according to funding agency requirements.

[Home](#) [Funding Agency Requirements](#) [How to Write a DMP](#) [Resources](#) [Research Data Management Symposium](#)

 Funding Agency Requirements

 Writing a DMP

 Resources

<https://guides.augusta.edu/DMP>

Data Trends



Data trends for the arts and humanities

- Digital Humanities – computational analysis on digitized text, objects
 - Data-mining/text-mining
 - topic modeling: discover topics within a collection of texts
 - Data visualization
- Spatial Humanities – geographical relationships of human culture
 - GIS/Story mapping



Biomedical Data Discovery

Data Discovery at the National Library of Medicine

As part of its work to fuel data-driven discovery and innovation, the National Library of Medicine has launched Data Discovery, a platform to provide access to datasets from selected NLM resources. Users can explore, filter, visualize, and export data in a variety of formats, including Excel, JSON, XML, as well as access and build with these datasets via API. NLM welcomes your comments on this program, as well as suggestions for datasets to be added to the catalog, at nlm-support@nlm.nih.gov.

Featured Content

Disaster Lit® 	TOXMAP - EPA Clean Air Markets - 2016 Coal Emissions 	Pillbox 
February 15, 2019 189 Views	September 21, 2018 126 Views	August 29, 2018 757 Views
		
(http://disasterlit.nlm.nih.gov) Disaster Lit®: Database for Disaster Medicine and Public Health is a database of links to disaster...	(https://toxmap.nlm.nih.gov/toxmap/) As a subset of TOXMAP this dataset represents 2016 coal emissions published by the US EPA's...	(https://pillbox.nlm.nih.gov) Pillbox contains metadata for oral solid dosage form medications, derived from FDA drug labeling...

7 Results

Sort by Alphabetical 

<h1>Dietary Supplements Label Database (DSLDD) - Ingredients</h1>		Drugs and Supplements	 Dataset
<p>(https://dsld.nlm.nih.gov) The Dietary Supplement Label Database (DSLDD) includes full label derived information from dietary supplement products marketed in the U.S. with a Web-based user interface that provides ready... More</p>		<p>Updated October 30, 2018</p> <p>Views 53</p>	
Tags supplements , labels		API Docs	

<https://datadiscovery.nlm.nih.gov/browse>



References

- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533, 452–454.
- Baillie, K. U. (2018). *Testing the reproducibility of social science research*.
<https://penntoday.upenn.edu/news/testing-reproducibility-social-science-research>
- Begley, C. G., & Ioannidis, J. P. (2015). Reproducibility in science: Improving the standard for basic and preclinical research. *Circ. Res.* 116, 116–126
- Buys, C. M., & Buys, S. (2015). Data management practices across an institution: Survey and report. *Journal of Librarianship and Scholarly Communication*, 3(2), p.eP1225. <http://doi.org/10.7710/2162-3309.1225>
- Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, 8(341). doi: 10.1126/scitranslmed.aaf5027
- Munafò, M. R., et al. (2017). A manifesto for reproducible science. *Nature Human Behavior*, 1, Article number: 0021.
<https://www.nature.com/articles/s41562-016-0021>
- Peels, R. (2019). Replicability and replication in the humanities. *Res Integr Peer Rev*, 4(2), DOI: 10.1186/s41073-018-0060-4.
- Stevens, J. R. (2017). Replicability and reproducibility in comparative psychology. *Front. Psychol.* <https://doi.org/10.3389/fpsyg.2017.00862>
- Thoegersen, J. L. (2018). Yeah, I guess that's data": Data practices and conceptions among humanities faculty. *portal: Libraries and the Academy*, 18(3), 491-504. DOI:10.1353/pla.2018.0030

****For a list of repositories discussed, please see handout.****

