

RESEARCH DATA SERVICES

Offering expert data assistance at every stage of the research process.

1: PLANNING

We can assist you with developing a data management plan and designing your planned data analysis, including:

- Implementing plans, using tools, and creating workflows for managing research data
- Advising on study design, power analysis, and choice of statistical methods
- Helping to meet increasingly stringent criteria from funding agencies

2: FINDING & COLLECTING

We have access to thousands of sources of data and experts who will help you:

- Locate, evaluate and format data
- Create metadata and data documentation protocols for new data collection
- Capture data using best practices and appropriate technology

3: ANALYZING

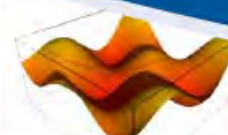
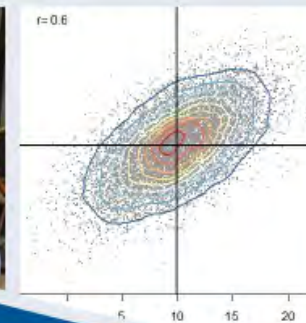
Get expert assistance from statistical, spatial, or media specialists to analyze your data and present your research:

- Learn to use cutting-edge tools and methods
- Experiment with high-resolution visualization technologies
- Develop graphical representations that bring impact to your analysis

4: SHARING & ARCHIVING

We can consult with you on strategies to help others discover or access your research by:

- Adhering to data sharing policies and norms
- Selecting a data-sharing repository
- Making your data easier to discover and reuse



Preserving and Sharing Data: Best Practices & Requirements for Selecting a Data Sharing Repository



<http://gpblog230.files.wordpress.com/2013/04/data-network.jpg>



<http://blogs.loc.gov/digitalpreservation/files/2012/02/filecabinet.jpg>

Bill Corey wtc2h@virginia.edu

Data Management Consulting Group
University of Virginia Library

What does “sharing data” mean?

Data sharing is the practice of making data used for scholarly research available to other investigators. Replication has a long history in science. The motto of The Royal Society is 'Nullius in verba', translated "Take no man's word for it." http://en.wikipedia.org/wiki/Data_sharing

Many funding agencies, institutions, and publishers have policies regarding data sharing because transparency and openness are considered by many to be part of the scientific method.

The logo for DataONE, featuring the word "Data" in blue, a globe icon, and "ONE" in blue.The logo for PLOS ONE, featuring a stylized atom icon, the word "PLOS" in bold black, and "ONE" in a smaller font.

Reasons to share my data...

- Receive scholarly credit
- Link to research products, create fuller picture
- Preserving data for future use
- Meet requirements (of funders, publishers)
- Some data are unique and cannot be replicated
- Other researchers can replicate and verify results
- Facilitate new discoveries
- Reduce duplication of data collection efforts
- Ease access to new researchers and citizen-scientists

What does “preserve data” mean?

Data preservation is the practice of managing data so it can be accessed and re-used by future researchers. Usually this term is applied to digital files (digital data preservation), but it is also important for analog objects.

“The goal of digital preservation is the accurate rendering of authenticated content over time.”

<http://www.ala.org/alcts/resources/preserv/defdigpres0408>

Preserving data helps to provide future access to the scientific, historical, and cultural heritage record of our world.

Selecting a data sharing repository

Questions to consider when selecting a repository or archive

Does your publisher specify a location for the data supporting an article?

Does your discipline recommend a specific repository or archive?

Does your funder identify a specific location or facility?

Does your institution have specific requirements?



http://datablog.is.ed.ac.uk/files/2013/12/bitsissue8_2-235x300.png

Selecting a data sharing repository

Best Practices

Choose early

Metadata

**Persistent
Identifiers**

Data embargo

Data access

Selecting a data sharing repository

Requirements

**What kinds of data
do they accept?**

**Do they require
specific file types?**

**Do they have a
rights or policy
statement?**

**Do they provide
long-term
preservation?**

Information potential users of your data will be looking for

When preparing to deposit your data in a repository, be sure to include this information in your documentation.

- Ownership of the data
- Permissions required for using the data
- Restrictions on reuse
- Documentation that makes it understandable and useful
- Data access – openly accessible, restricted access, or embargoed

Data Registries

Finding a repository can be confusing and time consuming.

International registry of data repositories:

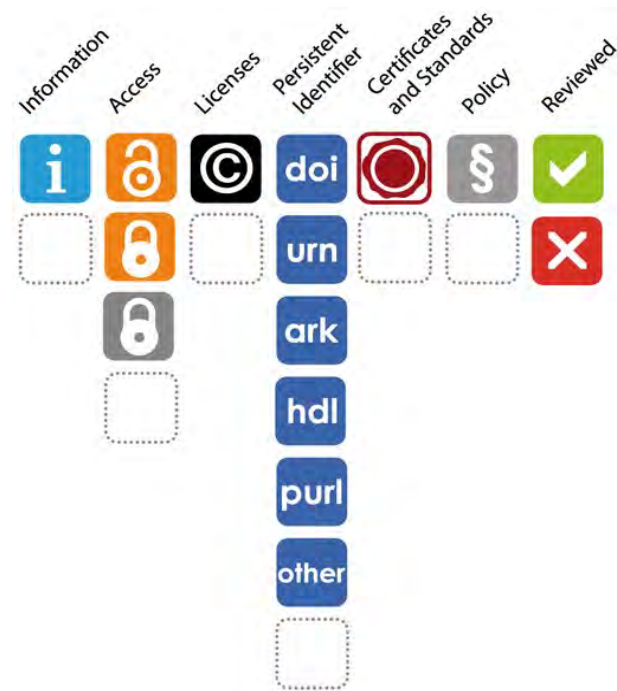
Re3data <http://www.re3data.org/>

Re3data merged with Databib in 2015 under the **DataCite** umbrella. All datasets will be provided with DOIs.

Advantages of Data Repository Registries

Registries provide information about the repositories that can be difficult to find.

- Do they provide Persistent Identifiers for finding and citing the data?
- Do they have access controls?
- What are their Terms of Use & Licenses?
- Are there guidelines for depositing data?
- Do they provide data preservation services?
- Do they have professional backup & documentation?
- What Repository Standards do they use?



<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0078080>

Data Registries exercise

Go to the registry:

Re3data <http://www.re3data.org/>

Identify your discipline or any that you are familiar with using the subject search, and see what options you can find. Some repositories will accept data, and some are simply data providers. What are your options?

Re3data icons

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

<http://re3data.org>

What do the icons mean?



http://www.plosone.org/article/fetchObject.action?uri=info:doi/10.1371/journal.pone.0078080.g001&representation=PNG_M

re3data.org

Record example

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

<http://re3data.org>

re3data is a global registry for research data repositories. Researchers can use it to select the appropriate repositories for the permanent storage and sharing of their data.

Repositories can be searched by name, country, subject, discipline, and content type. Clicking on the name will take you to a page with a description of the repository, including a URL.

The screenshot shows the 'Repository details' page for '3TU.Datacentrum'. The page includes a navigation menu with 'General', 'Institutions', 'Terms', and 'Standards'. The 'General information' section contains the following details:

Name of repository	3TU.Datacentrum
Additional name	3TU.DC 3TU.Datacentrum
Repository URL	http://datacentrum.3tu.nl/en/home/
Subjects	Engineering Sciences Natural Sciences
Description	3TU.Datacentrum is an archive for research data. It offers the knowledge, experience and the tools to share and safely store scientific research data in a standardized, secure and well-documented manner. 3TU.Datacentrum provides the research community with: • Advice and support on data management • A long-term archive for scientific research data • Support for current research projects • Tools for reusing research data.
Content types	Archived data Audiovisual data Images Plain text Raw data Scientific and statistical data formats Standard office documents Structured text
Keywords	technology and construction civil engineering aerospace engineering mechanical engineering electrical engineering materials engineering chemical technology applied natural sciences earth sciences design architecture product development biomedical technology
Repository size	5408 data sets
Repository type	institutional disciplinary
Mission statement for designated community	http://datacentrum.3tu.nl/en/about-3tudatacentrum/background-and-mission/
Research data repository language(s)	eng nld
Data and/or service provider	dataProvider

re3data.org

Record example

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

<http://re3data.org>

The **Institutions** tab lists the Responsible Institutions, including their responsibilities - funding, technical, or general -- their URL and contact information. The **Terms** tab includes links to Copyright policies, types of data access, and data deposit information. The **Standards** tab provides information about their repository, PI system used, if they are certified, and a general description.

The screenshot shows the 'Institutions' tab with two entries. The first entry is for TU Delft Library, and the second is for Technische Universiteit Eindhoven, Information Expertise Centre. The fields shown include Institution name, Additional name, URL, Contact, Country, Type(s) of responsibility, Type of institution, and Start date of responsibility.

Responsible institutions (4)	
Institution name	TU Delft Library
Additional name	Delft University of Technology Library
URL	http://www.library.tudelft.nl/en/
Contact	http://datacentrum.3tu.nl/en/about-Studatacentrum/contact/
Country	Netherlands
Type(s) of responsibility	funding general technical
Type of institution	non-profit
Start date of responsibility	2008
Institution name	Technische Universiteit Eindhoven, Information Expertise Centre
Additional name	
URL	
Contact	
Country	
Type(s) of responsibility	
Type of institution	
Start date of responsibility	
Institution name	
Additional name	
URL	
Contact	
Country	
Type(s) of responsibility	
Type of institution	
Start date of responsibility	
Institution name	
Additional name	
URL	
Contact	
Country	
Type(s) of responsibility	
Type of institution	
Start date of responsibility	

The screenshot shows the 'Terms' tab with sections for Legal aspects, Database access, Data access, Data licences, Data upload, and Application programming interfaces. The fields shown include Database access, Type of access to research, Data access, Type of access, License name, License URL, Data upload, Type of data, and Data upload r.

Terms (0)	
Legal aspects	
Database access	
Type of access to research	open
Data access	
Type of access	
License name	
License URL	
License name	
License URL	
License name	
License URL	
Data upload	
Type of data	
Data upload r	
Application programming interfaces (1)	
API type	API
URL	http://data.3tu.nl/repository/
Alerting services (1)	
Type of alerting service	RSS
Alerting service	http://datacentrum.3tu.nl/en/information/rss/
Remarks	
Remarks	3TU.Datacentrum is an initiative of the libraries of the three Dutch Technical Universities (TU Delft, TU Eindhoven and University of Twente). No policy yet: http://dataintelligence.3tu.nl/en/ii-data-management/data-policy/ All the data sets stored in the 3TU.Datacentrum are now accessible also via the NARCIS repository. NARCIS is the gateway to scholarly information in The Netherlands. Due to the link between NARCIS and 3TU.Datacentrum the research data sets can now be found even easier.
Entry date	2013-02-25
Last update	2013-08-23

The screenshot shows the 'Standards' tab with sections for Standards, Application programming interfaces, Alerting services, and Remarks. The fields shown include Name of the repository software, Persistent identifier system, Quality management, Certificates and Standards, API type, URL, Type of alerting service, Alerting service, and Remarks.

Standards	
Name of the repository software	other
Persistent identifier system	DOI
Quality management	yes
Certificates and Standards	DSA
Application programming interfaces (1)	
API type	API
URL	http://data.3tu.nl/repository/
Alerting services (1)	
Type of alerting service	RSS
Alerting service	http://datacentrum.3tu.nl/en/information/rss/
Remarks	
Remarks	3TU.Datacentrum is an initiative of the libraries of the three Dutch Technical Universities (TU Delft, TU Eindhoven and University of Twente). No policy yet: http://dataintelligence.3tu.nl/en/ii-data-management/data-policy/ All the data sets stored in the 3TU.Datacentrum are now accessible also via the NARCIS repository. NARCIS is the gateway to scholarly information in The Netherlands. Due to the link between NARCIS and 3TU.Datacentrum the research data sets can now be found even easier.
Entry date	2013-02-25
Last update	2013-08-23

Placing data in multiple locations

Having multiple locations to place your data, called data redundancy, is important when you are conducting your research so that you don't lose anything. Is it important when you are identifying a home for your research data after you have completed your project?

That will depend on your data. Is it something that can be collected again, or is it from a one-time event? Will other researchers be interested in using your data?

It doesn't hurt to place your data in multiple locations. This helps to ensure that it will be available to future users, and it increases the likelihood that others will be able to find it.

Consider placing your data in your institutional repository or archive

Libra

UVA Institutional Repository

- Opened in 2011
- Thesis and dissertations
- Articles
- Conference paper, posters
- Article preprint
- Book
- Chapter in an edited collection
- Datasets

The screenshot shows the LIBRA website interface. At the top, it says "LIBRA Online Archive of University of Virginia Scholarship". Below this is a search bar with the text "Search for open access materials and more." and a "Search" button. A note below the search bar reads: "Please Note: U.Va. users may [log in](#) to view items restricted to the University community." Below the search bar is a promotional message: "Libra makes publications available to the world and provides safe and secure storage for the scholarly output of the U.Va. community." To the right of this message are buttons for "Learn More", "Add Your Work", and a link for "Terms of Use".

On the left side, there is a "Limit results by:" section with the following categories and counts:

- Type of Work
 - [Article](#) (851)
 - [Doctoral Dissertation](#) (296)
 - [Master's Thesis](#) (111) [more »](#)
- Department or Academic Plan
 - [Department of Computer Science](#) (872)
 - [Department of Electrical and Computer Engineering](#) (38)
 - [Department of Systems Engineering](#) (37) [more »](#)
- Peer Reviewed** (36)

In the center, there is a photograph of the Rotunda at the University of Virginia.

On the right side, there are three sections:

- Open Access Works**: Faculty scholarly works available for research, scholarship, teaching and learning in a central, stable location. [Learn more...](#)
- Electronic Theses & Dissertations**: Find current theses and dissertations from departments and schools around Grounds. [Learn more...](#)
- Datasets**: Datasets may now be deposited into Libra. Additional features will become available in the future. Please send us your [feedback](#). [Learn more...](#)



- Provides/accepts data in several formats (ASCII, txt, SAS, SPSS, and Stata)
- Assigns a DOI (digital object identifier), links to persistent URL; facilitates data citation and location
- Creates variable-level DDI XML markup, makes data documentation machine-readable, allows more precise searching (e.g., by variable)
- Allows for online analysis of datasets
- Maintains secure data enclave, for archiving restricted data; accessible with permissions
- Data deposit form: <https://www.icpsr.umich.edu/cgi-bin/ddf2>
ICPSR becomes responsible for the management, cataloging and updating of data

openICPSR

Warning! This is a test site. Any material you upload, create, publish, or share will be periodically removed from our systems. Use at your own risk. All feedback is welcome, though, and can be submitted via the question mark at the bottom of the page.

Three easy steps: 1. Name your project 2. Upload and describe files 3. Purchase a curation option



[Watch our videos»](#)



Share your social and behavioral science research data

Get started now »

 **Maximize Access**

Be recognized and cited

 **Store Safely**

Store your data with confidence

 © Inter-university Consortium for Political and Social Research, 2014 [About](#) | [Contact](#)

ICPSR is creating a new repository that they will call openICPSR. It will be separate from the institutionally-funded ICPSR data repository, will require a fee to deposit the data, and will preserve for 10 years. In the Self-Deposit Package, the data depositor is responsible for the documentation & anonymization of the data.

Code & Multi-discipline Data Archives & Repositories

Open Science Framework: <https://openscienceframework.org/> Scientists can use OSF for free to archive, share, find, and register research materials and data.

figshare: <http://figshare.com/> “figshare is a repository where users can make all of their research outputs available in a citable, shareable and discoverable manner.”

GitHub: <https://github.com/> Share and archive code and collaborate with anyone.

runmycode: <http://www.runmycode.org/> Online repository allowing people to share and download computer code and data associated with scientific publications.

ResearchGate: <http://www.researchgate.net/> Online research social networking. Collaborate, share publications and data.

What if I can't find a home for my data?

Maybe your data doesn't 'fit' in a discipline-specific repository, and you don't feel comfortable putting it on a social network such as figshare or ResearchGate.

You can host it on a website, but there is an easier and better alternative:

The Dataverse Network



Roche DG, Lanfear R, Binning SA, Haff TM, et al. (2014) Troubleshooting Public Data Archiving: Suggestions to Increase Participation. PLoS Biol 12(1): e1001779. doi:10.1371/journal.pbio.1001779
<http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.1001779>

The Dataverse Network



A Web Application for Publishing, Citing,
Analyzing and Preserving Research Data



ABOUT GETTING STARTED BEST PRACTICES SOFTWARE APPS GUIDES



A repository for research data that takes care of **long term preservation** and **good archival practices**, while researchers can **share, keep control of** and get **recognition** for their data.

Supports the sharing of research data with a persistent **data citation**, and enables **reproducible research**.

Share & Find Data

The [Harvard Dataverse Network](#) is free* and open to all researchers worldwide to share, cite, reuse and archive research data.



Institutions or organizations may also choose to download the [open source software](#) for their own use. [Here](#) are some other [Dataverse Networks around the world](#).

* If you plan to upload >1TB please [contact us](#).

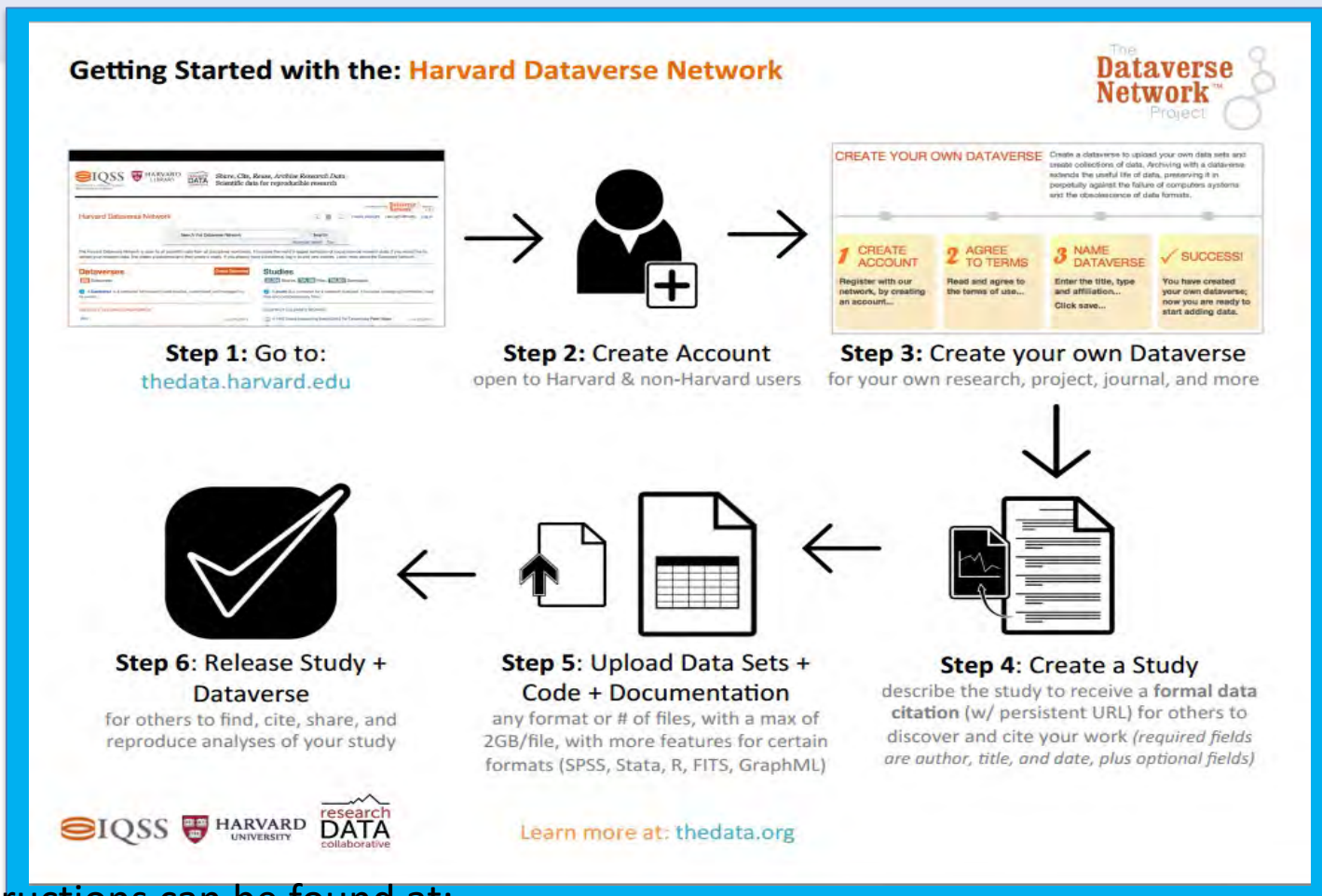


DVN hosts multiple, individually-branded Dataverses

- Researchers control the design, content, dissemination of their Dataverse, and can embed it in their own webpage
- DVN assigns handles (persistent id) and Universal Numerical Fingerprint (data fixity/verification)
- Extracts metadata for discovery, imports/exports metadata in multiple XML formats (DDI, Dublin Core, FGDC); data across DVNs searchable within one DVN
- Accepts data in multiple formats (Stata, SPSS, CSV), converts to preservation format
- Data can be subset, recoded, analyzed online
- Data sharing on DVN:

http://thedata.org/files/thedata_new2/files/gettingstartedguidefinal.pdf

Creating your own Dataverse



Specific instructions can be found at:

<http://thedata.harvard.edu/guides/dataverse-user-main.html>

Example of a Dataverse





Harvard Dataverse Network >

POWERED BY THE **Dataverse Network** PROJECT v. 3.6.2

Todd S. Sechser Dataverse

Search Studies [Advanced Search](#) [Tips](#)

Sort By: Studies: **4** | Downloads: **1259**

<p>Replication Data for: Crisis Bargaining and Nuclear Blackmail by Todd S. Sechser; Matthew Fuhrmann</p> <p>Description: Do nuclear weapons offer coercive advantages in international crisis bargaining? Almost seventy years into the nuclear age, we still lack a complete answer to this question. While scholars have devoted significant attention to questions ab... Continue [+]</p>	<p>hdl:1902.1/19713  116 downloads Last Released: Jan 22, 2013</p>
<p>Replication data for: Militarized Compellent Threats, 1918-2001 by Todd S. Sechser</p> <p>Description: The study of military coercion is a central topic in international relations, and in recent years research on coercive threats has yielded a long list of important theoretical innovations. In 1960, Thomas Schelling drew a distinction betwe... Continue [+]</p>	<p>hdl:1902.1/17315  253 downloads Last Released: Jan 3, 2013</p>
<p>Replication data for: The Illusion of Democratic Credibility by Alexander B. Downes; Todd S. Sechser</p> <p>Description: Do democracies make more effective coercive threats? An influential literature in international relations argues that democratic institutions allow leaders to credibly signal their resolve in crises, thereby making their threats more likel... Continue [+]</p>	<p>hdl:1902.1/18525  790 downloads Last Released: Aug 20, 2012</p>
<p>Replication data for: The Army You Have: The Determinants of Military Mechanization, 1979-2001 by Todd S. Sechser; Elizabeth N. Saunders</p> <p>Description: Recent research suggests that a crucial factor in understanding the outcomes of military conflicts is the extent to which militaries are mechanized — that is, their relative dependence on tanks and armored vehicles compared to manpower. Si... Continue [+]</p>	<p>hdl:1902.1/15670  100 downloads + analyses Last Released: Mar 21, 2011</p>

Example of a Dataverse record

Harvard Dataverse Network POWERED BY THE **Dataverse Network** PROJECT v. 3.6.2 SEARCH CREATE ACCOUNT LOG IN

RATLIFF & NOSEK (2010): CREATING DISTINCT IMPLICIT AND EXPLICIT ATTITUDES WITH AN ILLUSORY CORRELATION PARADIGM < View Previous Study Listing

hdl:1902.1/14569UNF:5:wwgNA3cCj3H9R+bcptOFca==
Version 2 - Released: Mon Jan 02 11:58:53 EST 2012

CATALOGING INFORMATION Data & Analysis Comments (0) Versions

Data Citation

If you use these data, please add the following citation to your scholarly references. Why cite?

Data Citation
Kate Ratliff; Brian Nosek, 2010, "Ratloff & Nosek (2010): Creating Distinct Implicit and Explicit Attitudes with an Illusory Correlation Paradigm", <http://hdl.handle.net/1902.1/14569UNF:5:wwgNA3cCj3H9R+bcptOFca>== Brian Nosek/Murray Research Archive

Citation Format Print

Publications
Ratloff, K. A., & Nosek, B. A. (2010). Creating Distinct Implicit and Explicit Attitudes with an Illusory Correlation Paradigm. *Experimental Social Psychology*, 46, 721-728.

Data Citation Details

Title	Ratloff & Nosek (2010): Creating distinct implicit and explicit attitudes with an illusory correlation paradigm
Study Global ID	hdl:1902.1/14569
Authors	Kate Ratliff (University of Tilburg); Brian Nosek (University of Virginia)
Producer	Brian Nosek, University of Virginia
Production Date	2010
Distributor	Brian Nosek, University of Virginia; Murray Research Archive
Contact	Kate Ratliff (University of Tilburg), kate.a.ratliff@gmail.com
Distribution Date	2010
Deposit Date	April 27, 2010
Original Dataverse	Brian Nosek Dataverse

Description and Scope

Description
Two studies used an illusory correlation procedure to test whether distinct learning episodes. All participants learned twice as much about the quality condition, the ratio of positive to negative information was equal between proportionally more positive qualities than the minority group, or vice versa conditions formed both implicit and explicit attitudes consistent with the condition showed the expected preference for the majority group (the illusory distinct influences on implicit and explicit attitude formation. The effects attitudes reflect accounting of covariation and explicit attitudes reflect in

Description Date 2010
Related Data <http://briannosek.com/>

Data Availability
Number of Files 2

Terms of Use
View Terms of Use [+]

Download the cataloging information in XML format - DDI (without variables) / DDI (full)

Harvard Dataverse Network POWERED BY THE **Dataverse Network** PROJECT v. 3.6.2 SEARCH CREATE ACCOUNT LOG IN

RATLIFF & NOSEK (2010): CREATING DISTINCT IMPLICIT AND EXPLICIT ATTITUDES WITH AN ILLUSORY CORRELATION PARADIGM < View Previous Study Listing

hdl:1902.1/14569UNF:5:wwgNA3cCj3H9R+bcptOFca==
Version 2 - Released: Mon Jan 02 11:58:53 EST 2012

Cataloging Information **DATA & ANALYSIS** Comments (0) Versions

Use the check boxes next to the file name to download multiple files. Data files will be downloaded in their default format. You can also download all the files in a category by checking the box next to the category name. You will be prompted to save a single archive file. Study files that have restricted access will not be downloaded.

Select all files Download Selected Files Total Number of Files: 2 Total Downloads: 42

<input type="checkbox"/>			
<input type="checkbox"/>	RN2010.Study1Data.tab Tab Delimited - 23 KB - 21 downloads + analyses	Download as... Access Analysis + Subsetting	"Clean" datafile for Study 1 View Data Citation [+]
	<input type="checkbox"/>	TABULAR DATA 104 Cases 41 Variables	
<input type="checkbox"/>	RN2010.Study2Data.tab Tab Delimited - 284 KB - 21 downloads + analyses	Download as... Access Analysis + Subsetting	"Clean" datafile for Study 2 View Data Citation [+]
	<input type="checkbox"/>	TABULAR DATA 893 Cases 61 Variables	

Data Sharing Archives & Repositories

Examples



Questions?

Thanks for attending today's workshop:

Preserving and Sharing Data: Best Practices & Requirements for Selecting a Data Sharing Repository

Please contact me if you have any questions, or need assistance in locating a data repository or preparing your data for sharing and archiving.

Bill Corey

Data Management Consultant

Research Data Services

University of Virginia Library

wtc2h@virginia.edu

We're available to help

- The UVA Library Research Data Services provides consulting and training services to UVA researchers and graduate students in all aspects of research data management.
- We can help you navigate and negotiate through the tricky issues and many approvals in order to responsibly share your research data.
- Contact us at dmconsult@virginia.edu



Photo credit
<http://vprompt.com/wp-content/uploads/2013/10/data-mining-300x154.jpg>