The race to preserve disappearing data
NSF Data Management Plan Requirements

Beginning January 18, 2011, proposals submitted to NSF must include a supplementary document of no more than two pages labeled "Data Management Plan" (DMP). This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. Proposals that do not include a DMP will not be able to be submitted. For more information about this new requirement, please see the Grant Proposal Guide, Chapter II.C.2.j and the Data Management and Sharing Frequently Asked Questions (FAQs).

Please note: the Engineering Directorate (ENG) has additional guidance for proposals submitted to ENG programs, [http://nsf.gov/eng/general/ENG_DMP_Policy.pdf](http://nsf.gov/eng/general/ENG_DMP_Policy.pdf). Questions and/or suggestions about this new requirement may be addressed to Dr. Maria K. Burka at mburka@nsf.gov.
3.2 Publication-related Research Data

**CIHR only**

Recipients of CIHR funding are required to adhere with the following responsibilities:

- Deposit bioinformatics, atomic, and molecular coordinate data into the appropriate public database (e.g. gene sequences deposited in GenBank) immediately upon publication of research results. Please refer to the [Annex](#) for examples of research outputs and the corresponding publicly accessible repository or database.
- Retain original data sets for a minimum of five years after the end of the grant.
Measures by measure: Diederik Stapel count rises again, to 54

with 6 comments

Diederik Stapel is up to 54 retractions.

Here’s the notice from Self and Identity:


The above named article has been retracted from publication in Self and Identity by agreement of the journal Editor, the Publisher and the Co-authors. The retraction has been agreed after the Noort Committee determined fraud in the publication. For further details, please visit the following link https://www.commissielevel.nl/

The paper has been cited just once, according to Thomson Scientific’s Web of Knowledge, in a paper published this spring called “Attractiveness of Limited Edition Artwork for First Generation Newly Affluent Consumers.” It’s Stapel’s second retraction from Self and Identity.
THE ADVENT OF MANDATORY DATA ARCHIVING

Daphne J. Fairbairn
Editor-in-Chief, Evolution

In the spring of 2010, *Evolution* joined with other key journals in evolution and ecology to adopt a formal policy requiring that data used in articles published in the journal be deposited in publicly accessible digital archives (Rauscher et al. 2010). This requirement was initially discretionary but becomes mandatory for all articles submitted on or after January 1, 2011. To signify this, the following paragraph has been added to our instructions for authors:

Data archiving. As a condition for publication, *Evolution* requires that data used in the paper are archived. DNA sequence data must be submitted to GenBank and phylogenetic data to TreeBASE. Other types of data must be deposited in an appropriate public archive such as Dryad, the NCEAS Data Repository or as supplementary online material associated with the paper published in *Evolution*. The data should be given with sufficient detail that, together with the contents of the paper, they allow each result in the published paper to be recreated. Authors may elect to have the data publicly available at time.

Any electronic format is acceptable for data archiving but widely used formats are clearly preferable. Comma-delimited text files (.txt) offer the widest accessibility and may become the norm for large datasets. Word, Excel, or even pdf formats may be adequate for smaller datasets or for files that contain a lot of text. However, text (.txt) files are likely to have a longer shelf-life because they should be readable in the future without proprietary software. In addition to the data files, authors must supply an accompanying text file explaining how the data files should be interpreted. The accompanying file should include helpful information such as expanded variable names, variable definitions, and details of the sampling design that may not be included in the published paper. To guard against use of the data in ways for which it is not appropriate, authors should note details such as whether the data are independent and collected at random if this was not specified in the associated paper.
Open Data Repositories

Recently published data


Latest from @datadryad

Dryad @datadryad

Attending #ISMTE2015 tomorrow? Be sure to join Meredith Morovati for a panel on "New Technologies to Support Editorial Office Operations."

Dryad @datadryad

Very excited to announce new integration w/Evidence-based Preclinical Medicine w/paid
Data Journals

Journal of open psychology data

Data Papers

Psychology data from the Race Implicit Association Test on the Project Implicit Demo website

Authors: Kaiyuan Xu, Brian Nosek, Anthony Greenwald

Abstract

This data archive includes Race Implicit Association Test (IAT) scores of 2,355,303 Internet volunteers who completed educational/demonstration versions of the Race IAT at https://implicit.harvard.edu from 2002 to 2012. Data in this archive can be downloaded for all years, either separately by year or in a single file. Codebooks, indicating the variable labels and value labels, and changes of variables over years, are available for both individual-year data sets and the entire data set. Participation in the (still on-going) Race IAT is interdisciplinary and open in the Race IAT Data Archive, also including a description of the...
Portage: Supporting Canadian innovation through shared expertise and stewardship of research data

Billions of dollars are invested every year in research, an investment that generates vast and diverse amounts of research data. If properly managed, these data have virtually limitless potential to be re-used in innovative ways. Sound research data management (RDM) practices, with due respect for confidentiality and intellectual property, accelerates scientific progress by allowing researchers to access and re-use others’ data for their own scientific purposes, thereby adding value to those data and speeding up the rate of new discoveries. It also leads to efficiencies by preventing duplication in data creation, and enables greater transparency and verification of research findings.

In Canada, this potential remains largely unrealized. Elsewhere, countries such as Australia, Germany, Netherlands, United Kingdom, and United States are investing in national policies, infrastructure and services to support more comprehensive RDM. As stated in the 2011 Report of the Canadian Research Data Summit, “Canada is one of the few advanced countries that does not yet have a national plan for managing the research data produced through public funding. As a result, valuable data are under-utilized and an important publicly funded asset is being wasted.”
Welcome.

DMP Builder is provided by the University of Alberta Libraries to help you write data management plans.

Learn about Data Management Plans
Looking to share your research data?
Find out more about the Libraries' research data services

Sign in
If you have an existing account.

Sign up
New to DMP Builder? Sign up today.

Please note that we are currently working on...
**Advanced Searching**

**Search Results** for Cataloging Information = "marital status"

**Original Dataverse**
ABACUS Public Data Collection (9)

**Author**
Ward, W. Peter (9)  
Gagné, Monique (9)

**Author Affiliation**

**Country/Nation**
United States (US) (3)  
United Kingdom (GB) (2)  
Ireland (IE) (1)  
Austria (AT) (1)  
Netherlands (NL) (1)

---

**Birth weight and economic growth data sets, Utrecht Hospital, 1880-1940, [2012]**  
by Gagné, Monique; Ward, W. Peter

Description: The variables contained in the data sets are primarily concerned with perinatal outcomes and maternal health. A number of variables with respect to the social and economic status of the mothers and their families were also included (ie. Oc...Continue [+]  

Production Date: 2012  
Producer: University of British Columbia Library. Data Services  
Distribution Date: May 25, 2012  
Related Material: http://hdl.handle.net/10573/42732

---

**Birth weight and economic growth data sets, The Rotunda (lying-in hospital), Dublin, 1869-1930, [2012]**  
by Gagné, Monique; Ward, W. Peter

Description: The variables contained in the data sets are primarily concerned with perinatal outcomes and maternal health. A number of variables with respect to the social and economic status of the mothers and their families were also included (ie. Oc...Continue [+]  

Production Date: 2012  
Producer: University of British Columbia Library. Data Services  
Distribution Date: May 25, 2012
### Subsetting and Recoding

#### Selected Variables
- aanmeldm
- aanmeldd
- delvrym
- delvryd
- year
- yearnr
- occupat

#### Recode & Case-Subset

1. Update the machine-generated new variable name and label in the input boxes. The new variable name must be unique within the current data file. For case-subsetting, the new variable becomes a condition variable that filters cases.

   - Condition Variable: occupat
   - New Variable Name: new_ui_occupat
   - New Variable Label: new_ui_Occupation - Notes

2. How to use the table for recoding: 📘
3. How to use the table for case-subsetting: 📘
4. How to enter a value or range as a condition: 📘

#### Drop

<table>
<thead>
<tr>
<th>Drop</th>
<th>occupat</th>
<th>New Value</th>
<th>New Value Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>musician</td>
<td>musician</td>
<td></td>
</tr>
<tr>
<td>☐</td>
<td>chairmende</td>
<td>chairmende</td>
<td></td>
</tr>
<tr>
<td>☐</td>
<td>strip artist</td>
<td>strip artist</td>
<td></td>
</tr>
<tr>
<td>☐</td>
<td>singer</td>
<td>singer</td>
<td></td>
</tr>
</tbody>
</table>
In-Browser Analytics

UBC Research Data Collection Dataverse

BOLOGNA SANT'ORSOLA MATERNITY CLINIC DATABASE, 1880-1940

Descriptive Statistics

Request ID: 055008
File Created: 2015-08-23 18:24:36.855 (US EST) - Note: will be erased one hour later.

MARSTAT: Marital status (code: 1=married; 2=unmarried; 3=widow)

<table>
<thead>
<tr>
<th>Value</th>
<th>Value Label</th>
<th>Freq</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Married</td>
<td>10593</td>
<td>85.3</td>
</tr>
<tr>
<td>2</td>
<td>Unmarried</td>
<td>1557</td>
<td>12.5</td>
</tr>
<tr>
<td>3</td>
<td>Widow</td>
<td>158</td>
<td>1.27</td>
</tr>
<tr>
<td>NA</td>
<td></td>
<td>109</td>
<td>0.878</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>12417</td>
<td>100</td>
</tr>
</tbody>
</table>

Median: 1 (Married)
Mode: 1 (Married)

BIRTHPL: Birthplace (code)

The number of categories is more than 50. Frequency/Percentage tables are not shown here

<table>
<thead>
<tr>
<th>Value</th>
<th>Value Label</th>
<th>Freq</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td></td>
<td>12417</td>
<td>100</td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Mode</td>
<td></td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>
Long Term Preservation

Preserving memory since 2009

Archivematica is a web- and standards-based, open-source application which allows your institution to preserve long-term access to trustworthy, authentic and reliable digital content.
Thanks!

lgoddard@uvic.ca