Data infused teaching and learning

Karen Visser
Analytical/statistical techniques

How real data is conceived, collected, analysed and managed

First class research output
-> first class teaching input
-> first class learning output
O₂ for UG T&L

- Global
- Accessible
- Collaborative
- Citable
- Published
DATA NEVER SLEEPS
How Much Data Is Generated Every Minute?

THE MOBILE WEB RECEIVES 217 NEW USERS.

WORDPRESS USERS PUBLISH 347 NEW BLOG POSTS.

571 NEW WEBSITES ARE CREATED.

FOURSQUARE USERS PERFORM 2,083 CHECK-INS.

FLICKR USERS ADD 3,125 NEW PHOTOS.

INSTAGRAM USERS SHARE 3,600 NEW PHOTOS.

APPLE RECEIVES ABOUT 47,000 APP DOWN LOADS.

TWITTER USERS SEND OVER 100,000 TWEETS.

EMAIL USERS SEND 204,166,667 MESSAGES.

FACEBOOK USERS SHARE 684,478 PIECES OF CONTENT.

CONSUMERS SPEND $272,070 ON WEB SHOPPING.

GOOGLE RECEIVES OVER 2,000,000 SEARCH QUERIES.

Flickr

YOUTUBE USERS UPLOAD 48 HOURS OF NEW VIDEO.

TUMBLR BLOG OWNERS PUBLISH 27,778 NEW POSTS.

APPLE RECEIVES ABOUT 47,000 APP DOWN LOADS.

BRANDS & ORGANIZATIONS RECEIVE 34,722 "LIKES."

Our Vision: More Australian researchers reusing research data more often

ANDS is enabling the transformation of:

<table>
<thead>
<tr>
<th>Data that are:</th>
<th>to</th>
<th>Structured Collections that are:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmanaged</td>
<td>→</td>
<td>Managed</td>
</tr>
<tr>
<td>Disconnected</td>
<td>→</td>
<td>Connected</td>
</tr>
<tr>
<td>Invisible</td>
<td>→</td>
<td>Findable</td>
</tr>
<tr>
<td>Single-use</td>
<td>→</td>
<td>Reusable</td>
</tr>
</tbody>
</table>

ands.org.au
Research Data Australia is a discovery service for Australian research data.

What's in Research Data Australia

- **Collections**
  - Research datasets or collections of research materials.
  - Browse All Collections (54,324)

- **Parties**
  - Researchers or research organisations that create or maintain research datasets or collections.
  - Browse All Parties (7,207)

- **Services**
  - Services that support the creation or use of research datasets or collections.
  - Browse All Services (140)

- **Activities**
  - Projects or programs that create research datasets or collections.
  - Browse All Activities (26,936)

Spotlight on research domains

More information on research data infrastructure for specific domains:

- TERN
- Australian Plant Phenomics Facility

**NCRIS and EIF Capability - Terrestrial Ecosystem Research Network**

TERN is a network of ecosystem scientists and infrastructure for the collection, cataloguing, storage and sharing of long-term ecosystem research data sets for science and management applications.
Who contributes to Research Data Australia

- AMMRF (12)
- ARC Centre of Excellence for Climate System Science (1)
- Atlas of Living Australia (484)
- AuScope (103)
- AusStage: Gateway to the Australian Performing Arts (1)
- AustLII (501)
- Australian Antarctic Data Centre (1,983)
- Australian Catholic University (15)
- Australian Institute of Health and Welfare (17)
- Australian Institute of Marine Science (371)
- Australian National Corpus (10)
- Australian Nuclear Science and Technology Organisation (192)
- Australian Ocean Data Network (14,777)
- Australian Synchrotron (3)
- BioGrid Australia Ltd (31)
- Bioplatforms Australia (1)
- Bond University (9)
- Breast Cancer Tissue Bank (18)
- Central Queensland University (30)
- Commonwealth Scientific and Industrial Research Organisation (478)
- Curtin University (20)
- Deakin University (98)
- Edith Cowan University (17)
- Flinders University (2)
- Geoscience Australia (14,165)
- Griffith University (43)
- Human Proteome Project (4)
- IMAS (240)
- James Cook University (1,633)
- La Trobe University (79)
- Macquarie University (1,068)
- Monash University (161)
- Museum Metadata Exchange (1,056)
- NICTA (3)
- OzFlux: Australian and New Zealand Flux Research and Monitoring (21)
- PARADISEC (177)
- Polar Information Commons (43)
- Publish My Data (46)
- QFAB (13,633)
- Queensland Department of Employment, Economic Development and Innovation (7)
- Queensland University of Technology (54)
- RMIT University (217)
- Southern Cross University (7)
- Swinburne University of Technology (11)
- Tasmanian Partnership for Advanced Computing (188)
- Terrestrial Ecosystem Research Network (648)
- The Australian National University (21)
- The University of Melbourne (1,018)
- The University of Newcastle, Australia (24)
- The University of Queensland (64)
- The University of Sydney (78)
- University of Adelaide (64)
- University of Ballarat (3)
- University of Canberra (5)
- University of New South Wales (162)
- University of Newcastle, Australia (3)
- University of South Australia (89)
- University of Southern Queensland (5)
- University of Technology, Sydney (4)
- University of the Sunshine Coast (7)
- University of Western Australia (29)
- University of Western Sydney (21)
- University of Wollongong (44)
- Victoria University (5)
A new opportunity

The Australian Curriculum places significant emphasis on the integration of ICT capabilities within science subjects. It states that Year 9 and 10 students will:

…employ their ICT capability to access information; collect, analyse and represent data; model and interpret concepts and relationships; and communicate science ideas, processes and information.
Data infused teaching

- Find it ready made
- Contribute to it
- Use it
Teaching with ESDS data

ESDS offers support for the secondary use of data across the research, teaching and learning communities. ESDS can help teachers on many aspects of using data resources in lectures and student dissertations, and for self-paced learning. The use of real-life data in teaching adds interest and relevance to courses, and, if the data are updated on a regular basis, ensures that the courses are pertinent to current substantive, theoretical and methodological debates.
Teaching with Data Simulations
helps students visualize abstract statistical concepts and see dynamic
processes behind the gathering, analysis, and interpretation of statistics.
Each sample activity includes instructions, teaching tips, assessment
ideas, and references. In one activity, students design an blind taste test
of two sodas. In another, students learn that larger sample sizes produce
better estimates and develop an appreciation for factors affecting
sampling variability. (SERC, National Science Foundation)

Interesting Fact:
The Coke vs. Pepsi Taste Test Challenge has students design and carry
out an experiment to determine whether or not students are able to
correctly identify two brands of cola in a blind taste test. This activity
allows students to gain a better understanding of the experimental
process and causality through considering control, random assignment,
and possible confounding variables.

Educational Attainment: A Data-Driven Learning Guide

Goal & Concept

Goal

The goal of this module is to explore the relationship between demographic characteristics and educational attainment. Frequencies and comparison of means will be used.

Concept

Educational attainment is a key concept for social scientists interested in social inequality, social mobility, and class. Educational attainment is closely related to occupation and income, with many better-paying jobs requiring a college degree or other advanced study. Thus, group differences in educational attainment are usually indicative of group differences in other areas of social life.

Examples of possible research questions about educational attainment:

- Do demographics like age, gender, race, nativity or disability relate to educational attainment, and if so how?
- What is the relationship between educational attainment and income?
- How does social inequality affect educational attainment?


This work is licensed under a Creative Commons Attribution-Noncommercial 3.0 United States License.
Data infused teaching

- Find it ready made
- Contribute to it
- Use it
.... provide opportunities for undergraduate students to participate in original research experiences within the framework of course curricula.

... pools student efforts across institutions to:
• improve DNA sequence quality
• generate handcurated gene models.

... students in that program:
• make significant gains in comprehension of topics in functional genomics
• show increased interest in research

By choosing regions of biological interest, so that the results contribute to research papers, as well as to the databases, much can be accomplished by a student-scientist partnership that otherwise might be difficult to attain.

SCIENCE 322 31 OCTOBER 10.1126/science.1165351
Data infused teaching

- Find it ready made
- Contribute to it
- Use it
Data infused learning

- Create it
- Understand it
- Visualise it
Redmap, a new and interactive website, invites the Australian community to spot, log and map marine species that are uncommon in Australia, or along particular parts of our coast.

Latest sighting:
Green moon wrasse
Spotted by David Harasti
Data infused learning

- Create it
- Understand it
- Visualise it
Quantitative Biology and Bioinformatics Specialisation

Biology needs Maths:
The vast quantities of genomic data generated by sequencing technologies can only be handled and interpreted through computers using algorithms underpinned by sound mathematical and statistical models.

Maths needs (biological) applications:
Mathematics also finds application in directly modelling biological processes at all levels from the molecular to macroscopic.

The intersection of mathematics and biology is one of the fastest growing areas of science, with most large biology laboratories now employing bioinformaticians, creating career opportunities in both the public and private sector.

http://studyat.anu.edu.au/2013/specialisations/
- Create it
- Understand it
- Visualise it
Hans Rosling's 200 Countries, 200 Years, 4 Minutes - The Joy of Stats - BBC Four

http://www.youtube.com/watch?v=jbkSRLYS0jo
A PERIODIC TABLE OF VISUALIZATION METHODS

Note: Depending on your location and connection speed it can take some time to load a pop-up picture.

© Ralph Lengler & Martin J. Eppler, www.visual-literacy.org

http://www.visual-literacy.org/periodic_table/periodic_table.html
Data infused libraries

- Include it
- Cite it
- Showcase it
Monash University is leading efforts to improve the management of research data. Well-managed research data is more discoverable and available for re-use, and...

http://monash.edu/library/researchdata/
The Online Learning Center at ICPSR offers Data-Driven Learning Guides that supplement undergraduate social science coursework with learning exercises using actual ICPSR data. They cover 42 subject areas in political science, sociology, and social psychology. They can be completed by the student or used for classroom demonstrations.

ICPSR has put a spotlight on the importance of statistical and quantitative literacy in recent years. To find out more about what they are doing to help faculty, visit their Teaching and Learning Web site. It provides information on their training activities but also links you to classroom-ready data and information about the Undergraduate Research Paper Competition.

TeachingWithData.org is another ICPSR-initiated project. The goal is to create a repository of educational materials designed to improve quantitative literacy skills in social science courses. Built especially for faculty teaching post-secondary courses in such areas as demography, economics, geography, political science, social psychology, and sociology, the materials include stand-alone learning activities, tools, and pedagogy services. The site is young (just launched in October 2009), but is set to be a very valuable teaching resource.

CensusScope.org is a portal to Census 2010 data presented in such a way that it can be understood by the novice user. Included are charts and trends, maps, population rankings and information on segregation.

DataCounts! is an interactive website designed by the Social Science Data Analysis Network to help integrate social statistics into the classroom setting. There are several collections of data included such as the American Community Survey Data from 2004-2006, the 1990 and 2000 censuses, census trend data from 1950-2000, the General Social Survey, and the Current Population Survey. Each collection contains a wide variety of datasets that can be integrated with WebCT, Blackboard, or other learning management systems.

http://guides.library.jhu.edu/content.php?pid=62430&sid=496594
Dataset Citations

Style Manual Examples

APA Style

APA 6th edition
For a complete description of citation guidelines refer to pp. 210-211 (dataset) and p. 212 (unpublished raw data) of the Publication Manual of the American Psychological Association, 6th edition [Call Number: Reference BF76.7.P63 2010].

Data set

Basic form:
Author/Rightsholder. (Year). Title of data set (Version number) [Description of form]. Location: Name of producer.
or
Author/Rightsholder. (Year). Title of data set (Version number) [Description of form]. Retrieved from http://

Example:

Unpublished raw data from study, untitled work

Basic form:

Example:
Data infused libraries

- Include it
- Cite it
- Showcase it
Last century data citation
John Gallant; Jenet Austin
(2012):
Contributing Area - Multiple Flow Direction (Partial) (3" resolution) derived from 1" SRTM DEM-H.
v1.
CSIRO. Data Collection.
http://dx.doi.org/10.4225/08/50A9D0E561DA6
Contributing Area - Multiple Flow Direction (Partial) (3" resolution) derived from 1" SRTM DEM-H

About this Data Collection

Data Collection Title: Contributing Area - Multiple Flow Direction (Partial) (3" resolution) derived from 1" SRTM DEM-H

Data Collection Description: CA_MFD_PARTIAL is contributing area in m2 computed using multiple flow directions on hillslopes and ANUDEM-derived flow directions in channels. The contributing area was computed on 1 degree tiles with 200 cell (about 5 km) overlaps so the areas in channels do not account for catchments beyond that size (hence the use of PARTIAL in the name). The p...

Field of Research: Environmental Management, Land Capability and Soil Degradation, Landscape Ecology, Natural Resource Management, Soil Sciences not elsewhere classified

DOI: 10.4225/08/50A9D0E561DA6

Data Start Date: 11 Feb 2000
Data End Date: 22 Feb 2000

Attribution Statement: John Gallant; Jenet Austin (2012): Contributing Area - Multiple Flow Direction (Partial) (3" resolution) derived from 1" SRTM DEM-H. v1. CSIRO. Data Collection. 10.4225/08/50A9D0E561DA6
Linking data and articles

Data from: Strong variations of mitochondrial mutation rate across mammals--the longevity hypothesis.

When using this data, please cite the original article:


Additionally, please cite the Dryad data package:

Data infused libraries

- Include it
- Cite it
- Showcase it
the NSF now allows for citable data (ie with a DOI) to be listed as an outcome of research, like a journal article. This is done in what is called a "biosketch" - basically a summary of your work, an a key part of the granting process.

Did you know?
NSF changed their rules for reporting your accomplishments.

You can now list products in your biographical sketch, not just publications.

"...including but not limited to publications, data sets, software, patents, and copyrights."

NSF Grant proposal guidelines Chapter II.C.2.f(i)(c)

Your data now counts as a product!

It needs to be both citable & accessible.
We can help you:

1. Archive and share your data in the Merritt repository: merritt.cdlib.org
2. Obtain a unique identifier from EZID: n2t.net/ezid
3. Start citing your data product!
Publishing frontiers: The library reboot

As scientific publishing moves to embrace open data, libraries and researchers are trying to keep up.

Richard Monastersky

27 March 2013

Sayeed Choudhury demonstrates the visualization wall, part of Johns Hopkins University’s drive to transform how its libraries and researchers deal with data.

WILL KIRK/JHU HOMewood PHOTOGRAPHY
AT THE FRONTLINE OF PUBLISHING IN SYSTEMATIC ZOOLOGY:
ZooKeys

IMPACT FACTOR 1.133

DEFINITION
ZooKeys is open-access, peer-reviewed, rapidly disseminated, online and print journal launched in July 2008 by Pensoft Publishers to accelerate research and free information exchange in taxonomy, phylogeny, and biogeography of animals (www.pensoftonline.net/zookeys).

SCOPE
New taxa descriptions, if accomplished with proper diagnoses, keys and/or revisions of at least species-group level; taxonomic revisions of extant (or "recent") and fossil animal groups; checklists and catalogues; phylogenetic and evolutionary analyses; papers in descriptive and/or historical biogeography; methodology papers; data mining and literature surveys; monographs, conspecti, atlases; collections of papers, Festschrift volumes, and conference proceedings.
GigaScience aims to revolutionize data dissemination, organization, understanding, and use. An online open-access open-data journal, we publish 'big-data' studies from the entire spectrum of life and biomedical sciences. To achieve our goals, the journal has a novel publication format: one that links standard manuscript publication with an extensive database that hosts all associated data and provides data analysis tools and cloud-computing resources.

Our scope covers not just 'omic' type data and the fields of high-throughput biology currently serviced by large public repositories, but also the growing range of more difficult-to-access data, such as imaging, neuroscience, ecology, cohort data, systems biology and other new types of large-scale sharable data.
Data infused LMSs

- Enable it

- Concept
- Capacity
- Capability
Curiosity driven learning

A/Pro Janette Lindesay
Associate Director (Education) & Deputy Director
The Fenner School of Environment & Society, ANU
### Data infused institutions

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 April</td>
<td>Research Data Licensing #2: Dr Kevin Cullen - Easy Access IP</td>
</tr>
<tr>
<td>23 April</td>
<td>Linking people to their research</td>
</tr>
<tr>
<td></td>
<td>Dr Laurel Haak - Executive Director, ORCID (USA)</td>
</tr>
<tr>
<td>30 April</td>
<td>Data Citation #1: DOIs and Data Citation</td>
</tr>
<tr>
<td></td>
<td>What you and your institution need to know</td>
</tr>
<tr>
<td>7 May</td>
<td>Data Citation #2: 2 DOI minting stories</td>
</tr>
<tr>
<td>21 May</td>
<td>Data Citation #3: Data Journals: a new publishing paradigm</td>
</tr>
<tr>
<td>4 June</td>
<td>Data Citation #4: What Griffith Uni are doing to establish a culture of data citation: a step by step approach</td>
</tr>
</tbody>
</table>

[ands.org.au/events](https://ands.org.au/events)
Questions, comments, chat?

Karen Visser
Program Leader, Resources, Policy & Skills

karen.visser@ands.org.au

http://creativecommons.org/licenses/by/4.0/