Research Data Management
Tools & Services

Daniel C. Tsang
FASRC Member
Data Librarian, UCI Libraries
Many academic libraries worldwide are now helping manage and provide stewardship for research data generated by faculty:

• Data are organized, elements identified, and the dataset is made retrievable

• Allows eventual deposit in secure repositories

• Facilitates subsequent data sharing
Data Sharing

Data Sharing occurs when, upon the agreement of a researcher, research data are made available to the research community for further, secondary analysis. Potential benefits include:

- Greater impact on the field
- Credit and citations
Participation in data sharing varies by discipline but recent federal mandates have placed the onus on researchers to come up with Data Management Plans indicating how you would manage your research data during your research project and beyond.

Your subject librarians at the UCI Libraries are here to help.
DMPTool

DMPTool, or Data Management Plan Tool, is a tool developed by the University of California Curation Center (UC3) at the California Digital Library (CDL) to help researchers prepare sponsored research grant applications that require data management plans.

Almost 3,000 plans have been drafted using this tool since Fall 2011.
DMPTool …

- Helps users create ready-to-use data management plans to meet specific funding agencies’ requirements
- Provides step-by-step instructions and guidance for how to manage data
- We encourage you to send the UCI Libraries a copy (to dtsang@uci.edu) of your data management plans so we can help, as in identifying potential data repositories
EZID is another CDL tool, to create & manage unique, long-term identifiers such as DOIs, for data files or digital objects. With EZID, one can:

- update current URL locations so citation links are never broken
- facilitate citations to a researcher’s data-related work giving it credit
UCI Libraries has licensed EZID so you can get DOIs for no extra charge. For more information, contact dtsang@uci.edu.

Over 246,000 DOIs have been allocated via EZID by CDL, which is a member of Data Cite.

DataCite offers a metadata search function that makes it easy to locate DOIs, including DOIs linked to datasets.
Data from: Policing effectiveness depends on relatedness and group size
 DOI: 10.5061/DRYAD.8042
 Walter, Bartosz Jerzy • Brunner, Elisabeth • Heinze, Jürgen
 title: Data from: Policing effectiveness depends on relatedness and group size

Crowd Dynamics, Policing and Hooliganism at Euro 2004
 DOI: 10.5255/UKDA-SN-5300-1 Dataset: Computer File
 Stott, C.
 title: Crowd Dynamics, Policing and Hooliganism at Euro 2004

Data from: Colony stage and not facultative policing explains pattern of worker reproduction in the Saxon wasp
 DOI: 10.5061/DRYAD.9PD1Q
 Bonckaert, Wim • van Zweden, Jelle S • d'Ettronde, Patrizia • Billen, Johan • Wenseleers, Tom
 title: Data from: Colony stage and not facultative policing explains pattern of worker reproduction in the subject: facultative worker policing
Policing_and_worker_fecundity_data

When using this data, please cite the original article:


Additionally, please cite the Dryad data package:


Dryad File Identifier: doi:10.5061/dryad.8042
9 downloads

Description
Worker policing, queen policing, workers and queens fecundity, survival of worker-laid eggs and queen-laid eggs according to colony size (large colony: 100 workers, small colony: 40 workers) and relatedness among workers in a colony (homogeneous colony: r = 0.75, heterogeneous colony: r = 0.1875). We observed three behavioral patterns: fertile workers were (1) policed by workers, (2) not policed but self-restrained from reproduction, or (3) non-policed and continued reproduction. We observed colonies where (1) all fertile workers were policed, (2) all fertile workers were not policed and continued reproduction, (3) all fertile workers were not policed but self-restrained from reproduction, and (4) mixture of policing, continued worker reproduction and self-restraint took place.

Keywords

Date Deposited
2010-12-06T20:30:46Z

Scientific Names
Temnothorax unifasciatus, Formicidae, Hymenoptera

Spatial Coverage
Garda Lake, Italy

Contained in Data Package
Data from: Policing effectiveness depends on relatedness and group size.

Show Full Metadata

Files in this Item

<table>
<thead>
<tr>
<th>Files</th>
<th>Size</th>
<th>Format</th>
<th>View</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policing_and_worker_fecundity_data.xls</td>
<td>33.28Kb</td>
<td>Microsoft Excel</td>
<td>View/Open</td>
<td>dataset-file</td>
</tr>
</tbody>
</table>

To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.
DataUp

An open source tool to help researchers document, manage, and archive tabular data

• Check for best practices
• Create metadata
• Get credit for data
• Archive & share data
The DataUp tool will parse your .xlsx or .csv file to detect the presence of potential issues that do not comply with data management best practices, such as embedded charts, embedded comments, columns with mixed data types.
Create Metadata

- DataUp will walk you through creating standard metadata using a form that becomes part of your spreadsheet, allowing for future use and sharing.
- When the data file is uploaded to a repository, the metadata generates a metadata file in Ecological Metadata Language (EML), which is a standard form of metadata for many Earth, environmental, and ecological researchers.
Once you have created metadata, you can connect directly to a repository via DataUp and upload your data for archiving.

Currently, DataUp is connected to the ONEShare, which is a dedicated DataUp repository that allows anyone to deposit tabular data.

ONEShare is a DataONE member node, which means datasets deposited there are indexed by DataONE and available for discovery and use by the public.
Merritt Repository

Merritt is a digital repository service that lets the University of California research community manage, archive, and share valuable digital content, including research data.
One example of a data service offered in collaboration with CDL via Merritt is DataShare: Open Data for the Global Scientific Community. The site makes available biomedical datasets from UCSF to the research community online. Some 300 publications have resulted from the data “published” online.
Web sites are constantly at risk of disappearing and the Web Archiving Service at CDL allows researchers to crawl and capture websites for research analysis.

At UCI Libraries we are also capturing regularly all web sites mounted on UCI public servers.
2007 Southern California Wildfires Web Archive

California Digital Library

Description
In October of 2007, a series of wildfires broke out throughout southern California. This web archive documents that event as it appeared on California State Agency sites, federal government sites, news, blogs, social networking sites.

In addition to providing lasting access to the webâ€™s coverage of the fires, this archive may serve as a point of comparison to other web archives of dramatic and quickly unfolding historical events. In particular it may show how people communicated about the fire, ...

more . . .

Search

Quick Facts
Sites: 152
Oldest site: 10/23/07
Most recent site: 11/06/07

Powered by the Web Archiving Service from the California Digital Library
See all web archives from the CDL
Materials in this web archive are archived copies for private study, scholarship and research.
Copyright © 2007-2013 The Regents Of The University of California.
UCISpace is UC Irvine Libraries’ digital repository for a variety of digital content including digital files, videos, audio recordings, photographs, text and other digitized content.

A current project: Collaborating with the School of Humanities to archive Vietnamese American oral histories.
UCI Libraries subject librarians are here to help at all stages of your research project, from at the beginning, during and after your research project has ended, throughout the life cycle of a dataset or research project, as you decide on how to document your research, apply metadata to research files and decide on the disposition of your research data.
Finding Repositories

At UCI Libraries, we have created a subject guide that not only links to the tools and services mentioned in this talk but also to discipline-based subject repositories such as those for digital humanities, social sciences and the sciences.

This resource will help you locate appropriate subject repositories as you develop your data management plans.
Data Management Tools

- Tools & Services from UC Curation Center
  These tools and services include the following:
  EZID - a service licensed to UCI researchers for identifying research files
  DMP Tool - a tool to help grant applicants fill out a concise data management plan geared to specific funding agencies. See tabs on this subject guide for more information.

Data Management & Sharing: Key Sites

Research Data is data generated in the research process. This subject guide delineates discipline-based research data repositories to help researchers and faculty seeking more information about potential repositories where they may be able to deposit their research data.

Careful Data Management of research data not only ensures preservation of such data but also helps in making such data shareable.

Sharing of research data contributes to scientific knowledge. Such data sharing has been the norm in some disciplines and is increasingly mandated by funding agencies, such as the National Science Foundation.

This guide is compiled by UCI Libraries’ bibliographers who serve on the Libraries’ Data Team, whose members are listed on the right column.

Listed below are some key resources:

Daniel C. Tsang
Research Data is data generated in the research process. This subject guide delineates discipline-based research data repositories that maintain and archive research data, to help researchers and faculty who seek more information about potential repositories where they may be able to deposit their research data.

- **DataCite**
  A compilation of data repositories through a collaboration between the British Library, BioMed Central and the Digital Curation Centre, that aims to capture the growing number of repositories for research data.

- **DataBib**
  An ongoing list of data repositories prepared by Purdue University Libraries. The list is arranged alphabetically and by subject.

- **3TU Datacentrum**
  Technical sciences, climate, fluid dynamics, earth sciences.

- **Center for International Earth Science Information Network**
  CIESIN works at the intersection of the social, natural, and information sciences, and specializes in on-line data and information management, spatial data integration and training, and interdisciplinary research related to human interactions in the environment.

- **DataBasin (US)**
  Centers are focal topics or geographies of special interest to Data Basin. Centers provide experts, datasets, maps, galleries, working groups that facilitate more effective collaboration.

- **Dryad (US/North Carolina)**
  Dryad is an international repository of data underlying peer-reviewed articles in the basic and applied biosciences. Dryad enables scientists to validate published findings, explore new analysis methodologies, repurpose data for research questions unanticipated by the original authors, and perform synthetic studies. Dryad is governed by a consortium of journals that collaboratively promote data archiving and ensure the sustainability of the repository.
Contacts

Send Data Management Plans & queries to:

Daniel C. Tsang, Data Librarian & FASRC member: dtsang@uci.edu

Your Subject Librarians: http://www.lib.uci.edu/about/contact/subject-librarians.html
• UCI Research Data Management Subject Guide: 
  http://libguides.lib.uci.edu/datamanagement (tools & data repositories)
• UCI NSF Data Management Plans: 
  http://nsfdmp.lib.uci.edu/
• UCISpace: http://ucispace.lib.uci.edu/
Resources: CDL

- UC Curation Center: http://www.cdlib.org/services/uc3
- UC Curation Center wiki: https://confluence.ucop.edu/display/Curation
- EZID: http://n2t.net/ezid
- Web Archiving Service: http://webarchives.cdlib.org
- Merritt: https://merritt.cdlib.org
- DMPTool: https://dmp.cdlib.org
- DataUp: http://dataup.cdlib.org
- DataPub Blog: http://datapub.cdlib.org
  (conversations about data)
- UCSF DataShare: http://datashare.ucsf.edu/xtf/search
  (example of open data site at Merritt)