La gestió FAIR de les dades de la recerca amb Dataverse

CSUC

Cicle de Conferències sobre la gestió de dades de recerca
Consorci de Serveis Universitaris de Catalunya
24 de Novembre, 2020

Mercè Crosas, Ph.D., Harvard University
scholar.harvard.edu/mercecrosas @mercecrosas

IQSS
The Institute for Quantitative Social Science

HARVARD UNIVERSITY

Photo credit: http://lovedevinandkelly.com/sightseeing/

FAIR everywhere

Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Sv
Velterop, Andra Waagmeester, Pet
-Show fewer authors

Scientific Data **3**, Article number: '

157k Accesses | **1991** Citations

Access & Citations

157k	1482	1991
Article Accesses	Web of Science	CrossRef

Online attention

- 1329 tweeters
- 6 Google+ users
- 1 F1000
- 3085 Mendeley
- 112 blogs
- 107 news outlets
- 1 Video uploaders
- 20 Citeulike
- 21 Facebook pages
- 1 Redditors
- 4 Wikipedia page

This article is in the 99th percentile (ranked 59th) of the 266,984 tracked articles of a similar age in all journals and the 1st percentile (ranked 1st) of the 1 tracked articles of a similar age in *Scientific Data*

- Published in 2016
- 54 authors
- > 1990 citations
- > 157K accesses
- Adoption by funding agencies, research communities

European Commission

Final Report and Action Plan from the European Commission Expert Group on FAIR Data

TURNING FAIR INTO REALITY

2018

Research and Innovation

European Commission

Six Recommendations for Implementation of FAIR Practice

By FAIR in Practice Task Force of the European Open Science Cloud FAIR Working Group

Independent Expert Report

EOSC Executive Board
WG FAIR
October 2020

Research and Innovation

National Health Institutes (NIH) FINAL Data Policy:

The National Institutes of Health (NIH) Policy for Data Management and Sharing (herein referred to as the DMS Policy) reinforces NIH's longstanding commitment to making the results and outputs of NIH-funded research available to the public through effective and efficient data management and data sharing practices. Data sharing enables researchers to rigorously test the validity of research findings,[5] strengthen analyses through combined datasets, reuse hard-to-generate data, and explore new frontiers of discovery. In addition, NIH emphasizes the importance of good data management practices, which provide the foundation for effective data sharing and improve the reproducibility and reliability of research findings. NIH encourages data management and data sharing practices consistent with the FAIR data principles.[6]

Final Supplemental Information: The final Supplemental Information asks researchers to describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools. This wording change is meant to highlight the importance of using a PID or other standard indexing tool so the data are findable, which is a key component of the FAIR (Findable, Accessible, Interoperable, and Reusable) Principles. PIDs are also listed as a desirable characteristic of data repositories in the Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research.

https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

Coalition for Publishing Data in Earth and Space Science Commitment Statement

To enable these principles

Repositories will strive to:

- Ensure that research outputs (e.g., data, software, technology, and physical samples) curated by repositories are open and FAIR, have essential documentation, and include human-readable and machine-readable metadata (e.g., on landing pages) in standard formats that are exposed and publicly discoverable.
- Ingest and expose data to promote interoperability and reuse.
- Ensure that unique, persistent identifiers are used for authors (e.g., ORCID), research objects (e.g., Digital Object Identifier), and physical samples (e.g., IGSN).
- Create associations among the research outputs that they manage and other related entities.
- Ensure that data and software have licenses that are as open as possible, and as protected as necessary.
- Support peer-review of related manuscripts by enabling access to the research outputs prior to publication.
- Gain third-party validation of trustworthy and sustainable practices and capabilities.

http://www.copdess.org/enabling-fair-data-project/commitment-to-enabling-fair-data-in-the-earth-space-and-environmental-sciences

15 FAIR PRINCIPLES

FINDABLE

4 principles

ACCESSIBLE

2 principles
2 sub-principles

INTEROPERABLE

3 principles

REUSABLE

1 principle
3 sub-principles

Emphasis on Machine-Actionability

" **The FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals.** "

Wilkinson et al. 2016. Nature-Springer Scientific Data. *The FAIR Guiding Principles for Scientific Data Management and Stewardship.* doi: 10.1038/sdata.2016.18

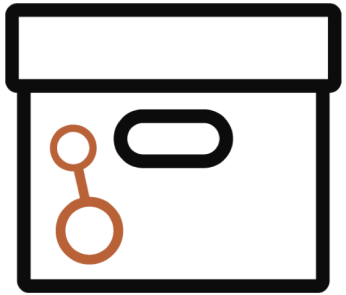
About Dataverse

Federated FAIR data repositories worldwide

- **Open-source**
- **63** installations
- **6** continents
- **7K** Dataverse collections
- **135K** datasets
- **800K** files
- **28M** file downloads

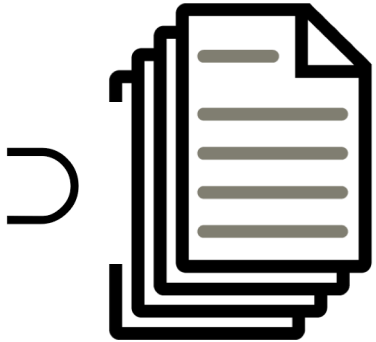
Developed at Harvard's Institute for Quantitative Social Science (IQSS) with contributions from the Dataverse community (https://dataverse.org)

Organization of a Dataverse Repository



Dataverse collection

Collection of datasets
Own administration
Own branding (& can be embedded in your site)



dataset

Citation
Metadata
Versioning
Terms/permissions
Collection of Files

File

Citation
Preview/Explore
Metadata
Versioning
Permissions

Key Dataverse Features

- ✓ Data Citation with **DOI for datasets and files** with credit to data authors
- ✓ Link from data to related article
- ✓ **Standard** schemas and custom metadata
- ✓ Access controls (open vs guestbook vs restricted) with **licenses and terms of use**
- ✓ Versioning and provenance
- ✓ **Descriptive Statistics** generated from **variables** in tabular data files
- ✓ Conversion to multiple formats of tabular data files
- ✓ Flexible upload of large data files (>> 5GB): Web UI, API, Standalone Client
- ✓ Integration with external tools through extensive **API**
- ✓ Data usage metrics with **Make Data Count**

Implementation of FAIR principles in Dataverse

Principle 1: To Be Findable

"Digital resources should be easy to find for both humans and computers. Extensive machine-actionable metadata are essential for automatic discovery of relevant datasets and services, and are therefore an essential component of the FAIRification process."

A. Jacobsen, et al. 2020. FAIR principles: Interpretations and implementation considerations. Data Intelligence 2(2020), 10-29. doi: 10.1162/dint_r_00024

Principle 1: To Be Findable

PRINCIPLE F1

(meta)data* are assigned a globally unique and persistent identifier

*(meta)data refers to data and metadata

DATAVERSE IMPLEMENTATION

- Support for DataCite DOIs; or Handles from Handle.net
- Always at the dataset level
- Optionally at the file level

Full, standard data citation automatically generated

HARVARD Dataverse

Add Data ▾ Search ▾ About User Guide Support Sign Up Log In

Replication Data for: To Emerge? Breadwinning, Motherhood, and Women's Decisions to Run for Office

Version 1.0

Teele, Dawn; Bernhard, Rachel; Shames, Shauna, 2020, "Replication Data for: To Emerge? Breadwinning, Motherhood, and Women's Decisions to Run for Office", https://doi.org/10.7910/DVN/S1EUAF, Harvard Dataverse, V1, UNF:6:gAlQI8fH9OpP/AdvZIo/1A== [fileUNF]

Cite Dataset ▾ Learn about Data Citation Standards.

- EndNote XML
- RIS
- BibTeX

Access Dataset ▾

Contact Owner | Share

Dataset Metrics

4 Downloads

Description anonymized replication data. (2020-09-28)

Subject Social Sciences

Keyword Candidate Emergence, American Politics, Gender an

Related Publication 202x. Bernhard, Rachel, Shauna Shames, and Dawn Women's Decisions to Run." Forthcoming: American

Data Citation, with **DataCite DOI**, fully compliant with Force11 **Joint Declaration of Data Citation Principles**

Principle 1: To Be Findable

PRINCIPLE F2

data are described with rich metadata

DATAVERSE IMPLEMENTATION

- Metadata standards in human- and machine-readable formats: Dublin Core; Documentation Data Initiative (DDI); DataCite; Schema.org
- Optional custom metadata

Support for multiple metadata standards

Files | Metadata | Terms | Versions

Export Metadata

- Dublin Core
- DDI
- DataCite
- DDI HTML Codebook
- JSON
- OAI_ORE
- OpenAIRE
- Schema.org JSON-LD

Citation Metadata

Dataset Persistent ID	doi:10.7910/DVN/S1EUAF
Publication Date	2020-10-14
Title	Replication Data for: To Emerge? Breadwinning, Motherhood, and Women's Decisions to Run for Office
Author	Teele, Dawn (University of Pennsylvania) - ORCID: 0000-0003-3079-3083 Bernhard, Rachel (UC Davis) Shames, Shauna
Contact	Use email button abov Teele, Dawn (University
Description	This is anonymized rep
Subject	Social Sciences
Keyword	Candidate Emergence,
Related Publication	202x. Bernhard, Rachel, Shauna Shames, and Dawn Teele. "To Emerge? Breadwinning and Income in Women's Decisions to Run." Forthcoming: American Political Science Review.
Depositor	Teele, Dawn
Deposit Date	2020-09-28

Rich support for Metadata Standards in **human- and machine-readable formats.**

Principle 1: To Be Findable

PRINCIPLE F3

metadata clearly and explicitly include the identifier of the data it describes

DATAVERSE IMPLEMENTATION

- ID is in the metadata tab of the Dataset landing page
- ID is in the metadata tab of the File landing page
- ID is included in exported metadata files

This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
▼<resource xmlns="http://datacite.org/schema/kernel-4" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://datacite.org/schema/kernel-4 http://schema.datacite.org/meta/kernel-4/metadata.xsd">
  <identifier identifierType="DOI">10.7910/DVN/94EU5T</identifier>
  ▼<creators>
    ▼<creator>
      <creatorName>Moundigbaye, Mantobaye</creatorName>
      <affiliation>(University of Canterbury)</affiliation>
    </creator>
    ▼<creator>
      <creatorName>Messemer, Clarisse</creatorName>
      <affiliation>(Bonneville Power Administration)</affiliation>
    </creator>
    ▼<creator>
      <creatorName>Parks, Richard W.</creatorName>
      <affiliation>(University of Washington)</affiliation>
    </creator>
    ▼<creator>
      <creatorName>Reed, W. Robert</creatorName>
      <nameIdentifier schemeURI="https://orcid.org/" nameIdentifierScheme="ORCID">0000-0002-6459-8174</nameIdentifier>
      <affiliation>(University of Canterbury)</affiliation>
    </creator>
  </creators>
  ▼<titles>
    ▼<title>
      REPLICATION DATA FOR: Bootstrap Methods for Inference in the Parks Model
    </title>
  </titles>
  <publisher>Harvard Dataverse</publisher>
  <publicationYear>2017</publicationYear>
  <resourceType resourceTypeGeneral="Dataset"/>
  ▼<relatedIdentifiers>
    <relatedIdentifier relatedIdentifierType="DOI" relationType="HasPart">doi:10.7910/DVN/94EU5T/LST9NX</relatedIdentifier>
    <relatedIdentifier relatedIdentifierType="DOI" relationType="HasPart">doi:10.7910/DVN/94EU5T/UNDYY5</relatedIdentifier>
    <relatedIdentifier relatedIdentifierType="DOI" relationType="HasPart">doi:10.7910/DVN/94EU5T/FGPKRO</relatedIdentifier>
    <relatedIdentifier relatedIdentifierType="DOI" relationType="HasPart">doi:10.7910/DVN/94EU5T/W24QCH</relatedIdentifier>
    <relatedIdentifier relatedIdentifierType="DOI" relationType="HasPart">doi:10.7910/DVN/94EU5T/VN0GOP</relatedIdentifier>
```


Dataset and Files DOIs in metadata

Principle 1: To Be Findable

PRINCIPLE F4

(meta)data are registered or indexed in a searchable resource

DATAVERSE IMPLEMENTATION

- DataCite metadata is registered and indexed by DataCite Search
- Schema.org metadata is indexed by Google Dataset Search

Google

Last updated | Download format | Usage rights | Topic | Free | Saved datasets

100+ datasets found

H

Data from: The International Political Economy Data...

dataverse.harvard.edu
search.datacite.org

zip, txt +4

Updated Aug 7, 2020

H

CCES Common Content, 2018

dataverse.harvard.edu
search.datacite.org

pdf, tsv +1

Updated Nov 17, 2019

H

Data from: Mining texts to efficiently generate global dat...

dataverse.harvard.edu

txt +1

Updated Jul 8, 2015

Data from: The International Political Economy Data Resource

Related Article

Explore at Harvard Dataverse | Explore at search.datacite.org

6 scholarly articles cite this dataset (View in Google Scholar)

zip, txt, application/gzip, text/tsv, type/x-r-syntax, docx

Unique identifier

https://doi.org/10.7910/DVN/X093TV

Dataset updated Aug 7, 2020

Dataset provided by

Harvard Dataverse

License

CC0 1.0 Universal Public Domain Dedication

License information was derived automatically

Description

Quantitative scholars in international relations often draw repeatedly on the same sources of country-year data across a diverse range of projects. The IPE Data Resource seeks to provide a public good to the field by standardizing and merging together variables from 89 IPE data sources into a single dataset, increasing

Principle 2: To Be Accessible

"Protocols for retrieving digital resources should be made explicit, for both humans and machines, including well-defined mechanisms to obtain authorization for access to protected data."

A. Jacobsen, et al. 2020.

Principle 2: To Be Accessible

PRINCIPLE A1

(meta)data are retrievable by their identifier using a standardized communications protocol

Sub-Principle A1.1: the protocol is open, free and universally implementable

Sub-Principle A1.2: the protocol allows for an authentication and authorization procedure, where necessary

DATAVERSE IMPLEMENTATION

- Support for HTTP (W3C), Rsync over ssh (GNU General Public license)
- RESTful API
- Authentication API Tokens
- Authorization service

API Guide

Contents: https://guides.dataverse.org

Principle 2: To Be Accessible

PRINCIPLE A2

metadata are accessible, even when the data are no longer available

DATAVERSE IMPLEMENTATION

- A deaccessioned dataset (data not available) is still findable and citable
- Metadata includes why the data are not available
- Metadata always accessible for restricted data files

Metadata always available

HARVARD Dataverse

Add Data ▾ Search ▾ About User Guide Support Sign Up Log In

2000 Utah Colleges Exit Poll

Deaccessioned

Contact Owner

David B. Magleby; Howard B. Christensen; Scott D. Grimshaw, 2019, "2000 Utah Colleges Exit Poll", https://doi.org/10.7910/DVN/2Z9KDF, Harvard Dataverse, V1, DEACCESSIONED VERSION, UNF:6:ME7YktcGved9FxnBuA4Ytw== [fileUNF]

Deaccession reason in dataset landing page when data not longer available

Deaccession Reason

User error. Do not use. Look under CSED and Utah Colleges Exit Poll

Versions

Dataset	Summary	Contributors	Published
1.0	Deaccessioned Reason: User error. Do not use. Look under CSED and Utah Colleges Exit Poll	CSED CSED	Dec 30, 2019

Access terms available for restricted data

HARVARD Dataverse

Add Data | Search | About | User Guide | Support | Sign Up | Log In

CAMEO Dataset: Detection and Prevention of "Multiple Account" Cheating in Massively Open Online Courses

Version 1.0

Northcutt, Curtis; Ho, Andrew; Chuang, Isaac, 2015, "CAMEO Dataset: Detection and Prevention of "Multiple Account" Cheating in Massively Open Online Courses", https://doi.org/10.7910/DVN/3UKVOR, Harvard Dataverse, V1

Cite Dataset | Learn about Data Citation Standards.

Contact Owner | Share

Dataset Metrics

Description

[NOTE: Data are currently only accessible to qualified reviewers.
For reviewers, detailed dataset descriptions are provided as text files associated with each dataset.]

This dataset includes statistics about student actions in MITx and HarvardX courses, used in an analysis of Copying Answers using Multiple Existences Online (CAMEO) behavior. The data are partially anonymized, but insufficiently so for open release. (2015-06-19)

Subject Computer and Information Science; Social Sciences; Other

Keyword CAMEO

Files | Metadata | Terms | Versions

Search this dataset... | Find

Filter by
File Type: All | Access: All | File Tag: All

Sort

1 to 4 of 4 Files

Request Access

cameo_candidate_master_harvester_pairs.csv
Plain Text - 634.0 KB - Jun 21, 2015 - 0 Downloads
MD5: 8f52e231fb316004c9668e65a6c7aa02
Please see "description_of_cameo_candidate_master_harvester_pair.txt"
CAMEO candidate master harvester pairs | Data

cameo_course_listings.csv
Plain Text - 8.0 KB - Jun 21, 2015 - 0 Downloads
MD5: 653639926109cc2dd0021db1cd7a3f14
Please see "description_of_cameo_course_listings.txt"
CAMEO course listings | Data

Optional **request access** feature for restricted data.
FAIR is not equal to Open.

Principle 3: To Be Interoperable

"When two or more digital resources are related to the same topic or entity, it should be possible for machines to merge the information into a richer, unified view of that entity. Similarly, when a digital entity is capable of being processed by an online service, a machine should be capable of automatically detecting this compliance and facilitating the interaction between the data and that tool."

A. Jacobsen, et al. 2020.

Principle 3: To Be Interoperable

PRINCIPLE I1

(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

DATAVERSE IMPLEMENTATION

- Linked data support with JSON-LD for Schema.org
- Data Documentation Initiative, DDI, (XML) as a rich schema to support extensive variable metadata
- DDI is unique providing detailed information for each variable

HARVARD Dataverse

Add Data | Search | About | User Guide | Support | Merce Crosas 3

ClimateRegressionData_150327.tab

Version 1.0

File Citation

Albouy, David, Graf, Walter, Kellogg, Ryan, and Wolff, Hendrik, 2018, "ClimateRegressionData_150327.tab", *Replication Data for: "Climate Amenities, Climate Change, and American Quality of Life" Journal of the Association of Environmental and Resource Economists 3, no. 1 (March 2016): 205-246.*, https://doi.org/10.7910/DVN/QCE1XY/BNJLIA, Harvard Dataverse, V1, UNF:6:CBIOoHJrG5/T6i+XjwBVwg== [fileUNF]

Cite Data File

Learn about Data Citation Standards.

File Metrics

45 Downloads

This file is part of "Replication Data for: "Climate Amenities, Climate Change, and American Quality of Life" Journal of the Association of Environmental and Resource Economists 3, no. 1 (March 2016): 205-246.".

Dataset Citation

Albouy, David, Graf, Walter, Kellogg, Ryan, and Wolff, Hendrik, 2018, "Replication Data for: "Climate Amenities, Climate Chang American Quality of Life" Journal of the Association of Environmental and Resource Economists 3, no. 1 (March 2016): 205-24 https://doi.org/10.7910/DVN/QCE1XY, Harvard Dataverse, V1, UNF:6:CBIOoHJrG5/T6i+XjwBVwg== [fileUNF]

Cite Dataset

Learn about Data Citation Standards.

Variable metadata from tabular data file

Preview | Metadata | Versions

Open View Data

	statefip	PumaID	msa	msaname	Wage_orig	Wage	Price
1	1	100100.0	2650	Florence, AL	-0.14469655	-0.15300082	-0.36732796
2	1	100200.0	3440	Huntsville, AL	-0.06367312	-0.0687066	-0.21142627
3	1	100300.0	3440	Huntsville, AL	-0.06052007	-0.06744661	-0.3109654
4	1	100400.0	19999	Non-metro, AL	-0.16140184	-0.166009	-0.49454302
5	1	100500.0	19999	Non-metro, AL	-0.16811557	-0.15688014	-0.40440822
6	1	100600.0	2030	Decatur, AL	-0.07162431	-0.09517802	-0.34356594
7	1	100700.0	19999	Non-metro, AL	-0.21245104	-0.19640322	-0.6055518

DDI supports interoperability even at the variable level

Data Explorer - Be

Español Françe

Replication Data for: "Climate Amenities, Climate Change, and American Quality of Life" Journal of the Association of Environmental and Resource Economists 3, no. 1 (March 2016): 205-246.

ClimateRegressionData_150327.tab

Albouy, David, Graf, Walter, Kellogg, Ryan, and Wolff, Hendrik, 2018, "Replication Data for: "Climate Amenities, Climate Change, and American Quality of Life" Journal of the Association of Environmental anc Resource Economists 3, no. 1 (March 2016): 205-246.", https://doi.org/10.7910/DVN/QCE1XY, Harvard Dataverse, V1, UNF:6:CBIOoHJrG5/T6i+XjwBVwg== [fileUNF]

1259 Results Download

18477636	msa	
18476854	msaname	
18476752	Wage_orig	Resiential-PUMA Wage Differential
18477053	Wage	Wage Differential based on Place of Wage
18476802	Price	Housing-cost differential
18477561	QOL_orig	QOL based on residential wage, no commuting
18477368	QOL_25_1	
18477175	QOL_GM	

First « 1 2 3 4 5 » Last

Records Per Page 10

Chart View Table View

Variable Price: Housing-cost differential

Values	Categories	N

Summary Statistics

Cases	N
	2057
	0
Maximum	1.4085395336151123
Minimum	-0.79858303
	-0.00509499
	-0.06162228
	0.35738088

Variable msaname:

Abilene,... 1

Extensive variable metadata (descriptive statistics) **automatically** derived from tabular data file in DDI format

Principle 3: To Be Interoperable

PRINCIPLE I2

(meta)data use vocabularies that follow FAIR principles

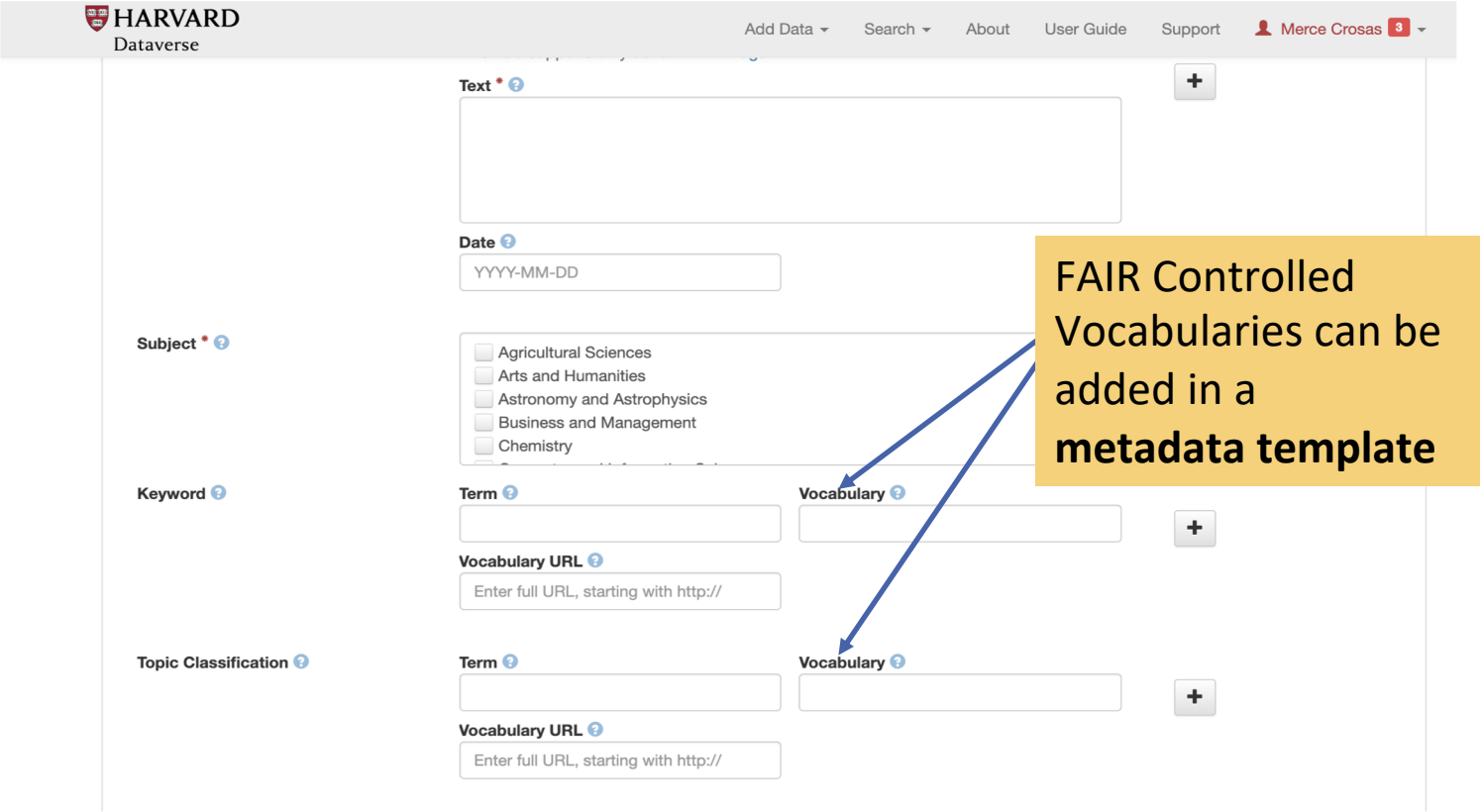
DATAVERSE IMPLEMENTATION

- FAIR controlled vocabularies and data models used in well-curated datasets
- Metadata template can help
- But, controlled vocabularies and ontologies **not** supported by default

Subject	Social Sciences
Topic Classification	mra 50 or fewer (Sampsize) http://www.murray.harvard.edu/vocabulary male (Gender) http://www.murray.harvard.edu/vocabulary 18-22 (Age) http://www.murray.harvard.edu/vocabulary White (Race) http://www.murray.harvard.edu/vocabulary student (SES) http://www.murray.harvard.edu/vocabulary 1 (Generations) http://www.murray.harvard.edu/vocabulary Special aspects of education (Education) http://authorities.loc.gov/ Mental health (Health) http://authorities.loc.gov/
Distributor	Murray Research Archive http://www.murray.harvard.edu
Distribution Date	1981
Time Period Covered	Start: 1930 ; End: 1940
Date of Collection	Start: 1930
Kind of Data	field study

Use **standard, global Controlled Vocabulary** from the Library of Congress

Enable use of standard vocabularies



Principle 3: To Be Interoperable

PRINCIPLE I3

(meta)data include qualified references to other (meta)data

DATAVERSE IMPLEMENTATION

- DDI schema supports references to other data
- Not yet supported: related objects in exported DataCite metadata

Australian National Political Attitudes, 1967: Supplemented with Treiman Prestige Scores (M023V1)

Version 2.0

Donald Treiman, 2012, "Australian National Political Attitudes, 1967: Supplemented with Treiman Prestige Scores (M023V1)", https://doi.org/10.7910/DVN/D1NDDL, Harvard Dataverse, V2

Cite Dataset

Learn about Data Citation Standards.

Dataset Metrics

2 Downloads

Related Material

McDonnell, Patrick, Leonard Blom, F. Lancaster Jones, and Paul Duncan-Jones, "Notes on the Australian Occupational Classification," Australian National University, Paper preared fro annual meeting of the Sociological Association of Australia and New Zealand, August 1976. Australia, Bureau of Census and Statistics, "Classification and Classified List of Occupations" (revised June 1961), Government Printer, Canberra, 1961. Australia, Bureau of census and Statistics, "Index of Occupations" (revised June 1961), Government Printer, Canberra, 1961. Broom, Leonard, F. Lancaster Jones and Jerzy Zubrzycki, "A Occupational Classification of the Australian Workforce," THE AUSTRALIAN AND NEW ZEALAND JOURNAL OF SOCIOLOGY Vol. 1, No. 2 (October, 1965), p.1-2.

Related Datasets

Aitkin, Donald, Michael Kahan, and Donald E. Stokes. AUSTRALIAN NATIONAL POLITICAL ATTITUDES, 1967. Conducted by Donald Aitkin and Michael Kahan, Australian National University, and Donald E. Stokes, University of Michigan. ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], 1975. doi:10.3886/ICPSR07282.v1; Aitkin, Donald, Michael Kahan, and Donald E. Stokes. Australian National Political Attitudes, 1969. ICPSR07393-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2001. doi:10.3886/ICPSR07393.v1

Reference in metadata to **related datasets** or other research objects

Principle 4: To Be Reusable

"Digital resources are sufficiently well described for both humans and computers, such that a machine is capable of deciding: if a digital resource should be reused; if a digital resource can be reused, and under what conditions; and who to credit if it is reused."

A. Jacobsen, et al. 2020.

Principle 4: To Be Reusable

PRINCIPLE R1

(meta)data are richly described with a plurality of accurate and relevant attributes

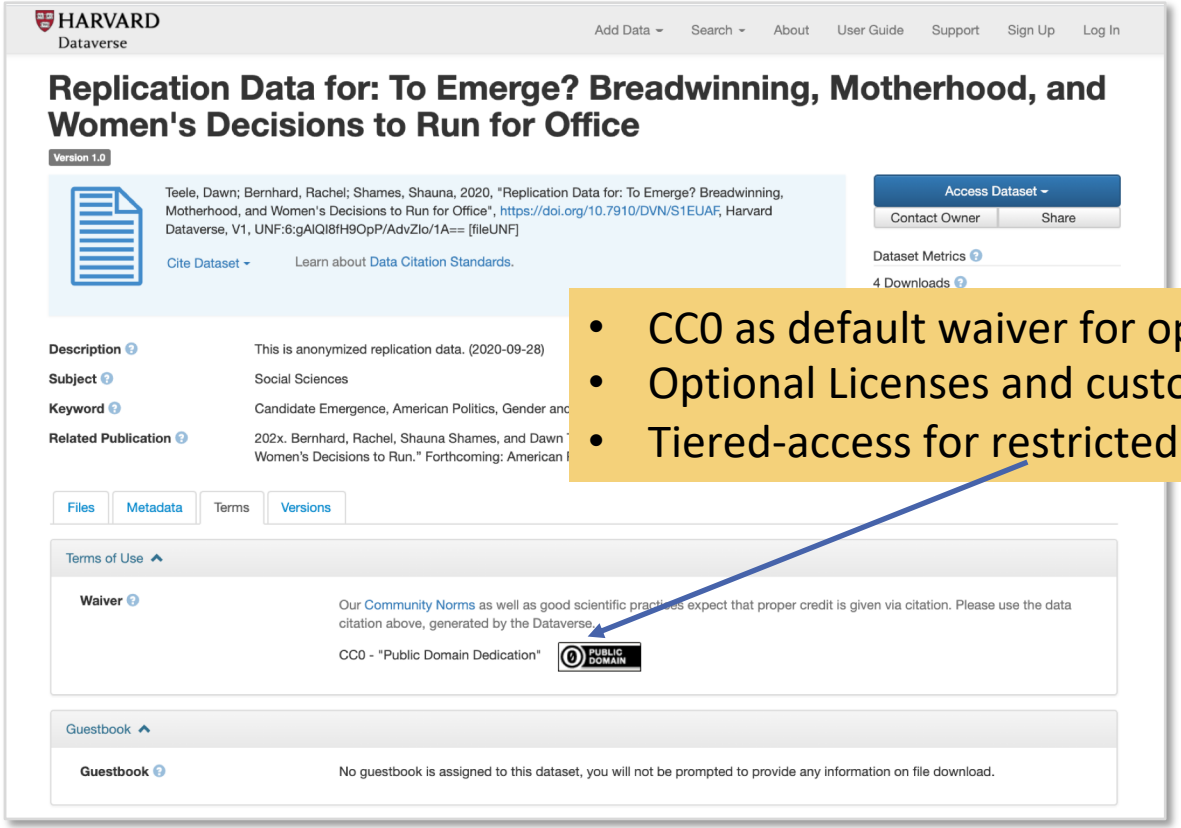
Sub-Principle R1.1:
(meta)data are released with a clear and accessible data usage license

DATAVERSE IMPLEMENTATION

Included in Metadata:

- Data use license/waiver
- Data access and use terms

Licenses, Terms, and Tiered-Access to Data



- CC0 as default waiver for open data;
- Optional Licenses and custom Terms;
- Tiered-access for restricted data

Principle 4: To Be Reusable

PRINCIPLE R1

Sub-Principle R1.2:
(meta)data are associated with detailed provenance

DATAVERSE IMPLEMENTATION

- Full data citation metadata with credit to data authors, providers, distributors
- Versions with changes documented automatically
- W3C PROV support

HARVARD Dataverse

Add Data | Search | About | User Guide | Support | Sign Up | Log In

Files | Metadata | Terms | Versions

View Differences

	Dataset	Summary	Contributors	Published
	8.0	**Files (Added: 2; Removed: 2);** View Details	Sergio Petralia	Jan 19, 2019
	7.2	**Terms of Use/Access Changed** View Details	Sergio Petralia	Dec 19, 2017
	7.1	**Citation Metadata:** Author (3 Changed); Contact (1 Changed); **Additional Citation Metadata:** (5 Changed); View Details	Sergio Petralia	Oct 4, 2017
	7.0	**Files (Added: 1; Removed: 1; Replaced: 1);** View Details	Sergio Petralia	Sep 12, 2017
	6.1	**Citation Metadata:** Notes (Changed); View Details	Sergio Petralia	Jan 10, 2017
	6.0	**Files (Added: 2; Removed: 2);** View Details	Sergio Petralia	Jan 10, 2017
	5.0	**Files (Added: 1; Removed: 1);** View Details	Sergio Petralia	Oct 3, 2016
	4.0	**Citation Metadata:** Description (1 Changed); **Additional Citation Metadata:** (1 Added, 8 Changed); **Files (Added: 2; Removed: 2; Changed File Metadata: 1);** View Details	Sergio Petralia	Sep 29, 2016
	3.3	**Additional Citation Metadata:** (5 Changed); View Details	Sergio Petralia	Sep 7, 2016
	3.2	**Additional Citation Metadata:** (2 Added); View Details	Sergio Petralia	Sep 7, 2016
	3.1	**Additional Citation Metadata:** (5 Added); View Details	Sergio Petralia	Sep 7, 2016
	3.0	**Files (Added: 1; Removed: 1);** View Details	Sergio Petralia	Sep 7, 2016
	2.1	**Citation Metadata:** Notes (Changed); View Details	Sergio Petralia	Sep 1, 2016
	2.0	**Files (Added: 1);** View Details	Sergio Petralia	Sep 1, 2016
	1.1	**Additional Citation Metadata:** (2 Added); View Details	Sergio Petralia	Sep 1, 2016
	1.0	This is the first published version.	Sergio Petralia	Aug 26, 2016

Principle 4: To Be Reusable

PRINCIPLE R1

Sub-Principle R1.3:
(meta)data meet domain-relevant community standards

DATAVERSE IMPLEMENTATION

- DDI for social science data
- FITS for astronomy data
- Metadata blocks for other community standards
- File format conversion to reusable formats (tabular)

HARVARD Dataverse

Add Data | Search | About | User Guide | Support | Sign Up | Log In

Cepheus-L1251 Data

Version 1.0

Keown, Jared, 2017, "Cepheus-L1251 Data", https://doi.org/10.7910/DVN/SD8QCL, Harvard Dataverse, V1

Cite Dataset

Learn about Data Citation Standards.

Dataset Metrics

152 Downloads

Description Green Bank Ammonia Survey data for Cepheus-L1251

Subject Astronomy and Astrophysics

Files | Metadata | Terms | Versions

Metadata extracted from Astronomy FITS files

Search this dataset... Find

Filter by

File Type: All Access: All

Sort

1 to 8 of 8 Files Download

Cepheus_L1251_C2S_cube.fits
FITS - 305.9 MB - Oct 16, 2017 - 26 Downloads
MD5: 5e1e1154718aba67b9674fbed1feec41
This is a FITS file with 1 (primary) HDU. The following recognized metadata keys have been found in the FITS file: CRVAL2; NAXIS2; NAXIS; INSTRUME; NAXIS1; NAXIS0; TELESCOP; CRVAL1;

Download

Cepheus_L1251_C2S_params.fits
FITS - 3.7 MB - Oct 16, 2017 - 14 Downloads
MD5: efedd1f0d94c85b2a2c9bfefe10f004c
This is a FITS file with 1 (primary) HDU. The following recognized metadata keys have been found in the FITS file: CRVAL2; NAXIS2; NAXIS; INSTRUME; NAXIS1; NAXIS0; TELESCOP; CRVAL1;

Download

Conclusions

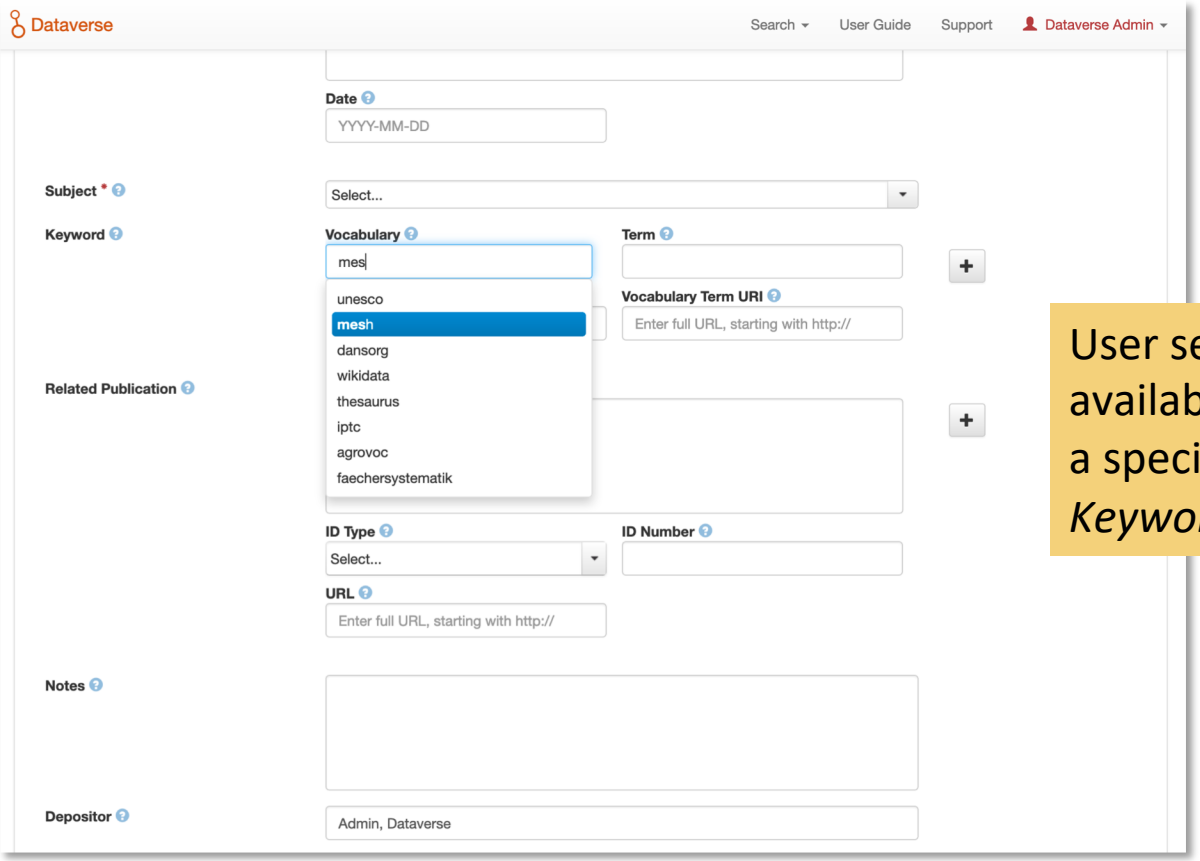
FM [AID*]	Question	Dataverse Q'aire	Dataverse Optimized
Identifier type	1	DOI	DOI
F1A	2		
F1B	Not tested in Q'aire		
F2A	4A		
F2A	4B		
F3	5B		
F4	6A	FIXED	
F4	6B		
A1.1	7A		
A1.2	8A		
A1.2	8B	N/A	N/A
A2	9		
I1	10		
I2	11		
I3	12		
R1.1	13		
R1.2	14A		

FAIR AUTOMATED TEST

M. Wilkinson, et al. 2019. Evaluating FAIR-Compliance Through an Objective, Automated, Community-Governed Framework. Nature Scientific Data, Issue 6, Article 174

- Dataverse scored **well in Findable, Accessible, and Reusable** principles
- Since this FAIR test (in 2018), all Findable principles have been implemented in Dataverse
- **Interoperable** principles support is now **under improvement**

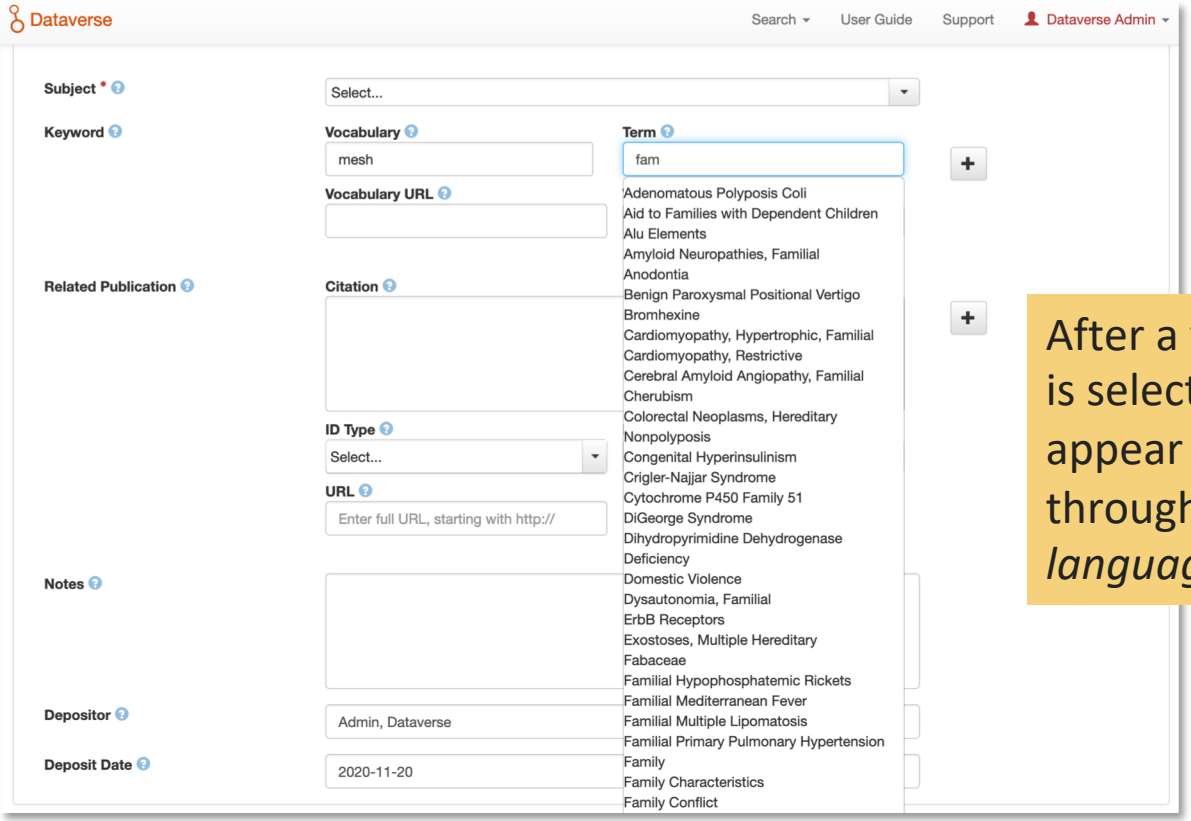
Improving Interoperable Principles: Controlled Vocabularies



User selects from the available list of *ontologies* for a specific field, for example, *Keyword* (configurable)

Contribution by Vyacheslav Tikhonov under review

Improving Interoperable Principles: Controlled Vocabularies



After a vocabulary for *Keyword* is selected, suggested *terms* will appear (information obtained through external API, per *language* and *vocabulary*)

Contribution by Vyacheslav Tikhonov under review

Improving Interoperable Principles: Controlled Vocabularies

Dataverse

Search User Guide Support Dataverse Admin

Subject *
Select...

Keyword
Vocabulary: mesh
Term: Familial Hypophosphatemic Rickets
Vocabulary URL: http://www.yso.fi/onto/mesh/
Vocabulary Term URI: http://www.yso.fi/onto/mesh/D053098

Related Publication
Citation
ID Type: Select...
ID Number
URL: Enter full URL, starting with http://

Notes

Depositor: Admin, Dataverse

Deposit Date: 2020-11-20

The *Vocabulary Term URI* is filled automatically upon *Term* selection

Contribution by Vyacheslav Tikhonov under review

Summary

The Dataverse software:

- ✓ Provides strong support for FAIR principles
- ✓ Is adding standard vocabularies and ontologies to improve interoperability

Software support is not enough and there is always room for improvement:

- ✓ FAIR principles are not a standard, they are guidelines
- ✓ Support for FAIR is a process, it can always be improved
- ✓ FAIR data enables the use of automated tools and AI algorithms on the data
- ✓ Software support is not everything, **data curation** remains important
- ✓ Beyond FAIR, we need to consider **data quality** and **responsible data sharing**

Gràcies

Mercè Crosas, Ph.D., Harvard University
scholar.harvard.edu/mercecrosas @mercecrosas

IQSS
The Institute for Quantitative Social Science

HARVARD UNIVERSITY