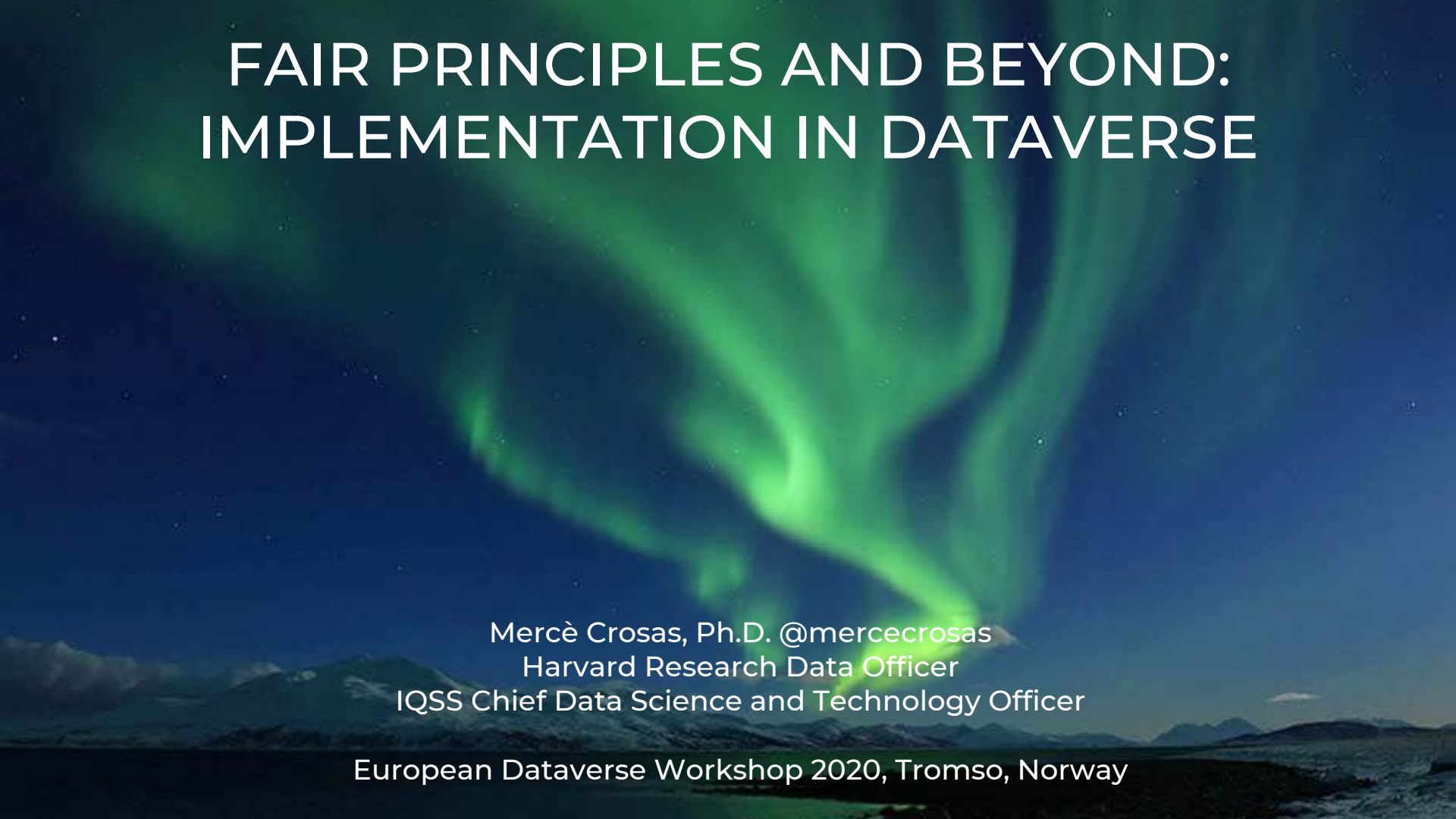
FAIR PRINCIPLES AND BEYOND: IMPLEMENTATION IN DATAVERSE



Mercè Crosas, Ph.D. @mercecrosas
Harvard Research Data Officer
IQSS Chief Data Science and Technology Officer

European Dataverse Workshop 2020, Tromso, Norway

THIS TALK

FAIR Principles

- The importance of FAIR
- Implementation in Dataverse

Beyond FAIR:

- Responsible FAIR
- Data Quality
- Reproducibility

FAIR PRINCIPLES BECOME POPULAR

SCIENTIFIC DATA

Comment | 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. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons - Show fewer authors

- Published in 2016
- 54 authors
- About 1000 citations
- 85K accesses
- Altmetric: Ranked 64th of 265K articles of same age

FAIR EVERYWHERE

European Commission Data Policy:

"**The OBJECTIVES for adopting and implementing the JRC (Joint Research Centre) Data Policy include: … Facilitate management, broaden access and use of the JRC data, in line with FAIR Data principles (Findable, Accessible, Interoperable, Reusable);**"

https://ec.europa.eu/jrc/en/publication/jrc-data-policy

FAIR EVERYWHERE

National Health Institutes (NIH) proposed DRAFT Data Policy:

"NIH encourages data management and data sharing practices consistent with the NIH Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research and the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles."

https://osp.od.nih.gov/wp-content/uploads/Draft_NIH_Policy_Data_Management_and_Sharing.pdf

FAIR EVERYWHERE

Coalition for Publishing Data in Earth and Space Science (CODPESS) Commitment Statement:

"Ensuring that Earth, space, and environmental science research outputs, including data, software, and samples or standard information about them, are open, FAIR, and curated in trusted domain repositories whenever possible ..."

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

But, what does it mean?

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.*

But, what does it really mean?

Step by step implementation in Dataverse

DATAVERSE REPOSITORIES

dataverse **dataset** **File**

15 FAIR PRINCIPLES

FINDABLE

4 principles

ACCESSIBLE

2 principles,
2 sub-principles

INTEROPERABLE

3 principles

REUSABLE

1 principle,
3 sub-principles

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."

Jacobsen et al. 2016. *FAIR Principles: Interpretations and Implementation Considerations(Forthcoming)*

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

REPLICATION DATA FOR: Bootstrap Methods for Inference in the Parks Model

Version 3.0

Moundigbaye, Mantobaye; Messemer, Clarisse; Parks, Richard W.; Reed, W. Robert, 2017, "REPLICATION DATA FOR: Bootstrap Methods for Inference in the Parks Model", https://doi.org/10.7910/DVN/94EU5T, Harvard Dataverse, V3

Cite Dataset

- EndNote XML
- RIS
- BibTeX

Learn about Data Citation Standards.

Dataset Metrics

126 Downloads

This dataset contains all the materials needed to reproduce the results in "Bootstrap Methods for Inference in the Parks Model". Please read the README document first. The results were obtained using SAS/IML software, and the files consist of SAS data sets and SAS programs. (2019-06-06)

Subject Business and Management; Social Sciences

Keyword Parks model, PCSE estimator, SUR (Seemingly Unrelated Regression), Panel data, Bootstrap, Cross-sectional dependence

DataCite DOI in standard Data Citation, compliant with Joint Declaration of Data Citation Principles in addition to F1

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
- Standard metadata for data citation, collection, design, and analysis, variables (DDI)
- Optional custom metadata

Files | Metadata | Terms | Versions

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

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/94EU5T
Publication Date	2017-10-26
Title	REPLICATION DATA FOR: Bootstrap Methods for Inference in the Parks Model
Author	Moundigbaye, Mantobaye (University of Canterbury) Messemer, Clarisse (Bonneville Power Administration) Parks, Richard W. (University of Washington) Reed, W. Robert (University of Canterbury) - ORCID: 0000-0002-6459-8174
Contact	Use email button above to contact. Reed, W. Robert (University of Canterbury)
Description	This dataset contains all the materials needed to reproduce the results in "Bootstrap Methods for Inference in the Parks Model". Please read the README document first. The results were obtained using SAS/IML software, and the files consist of SAS data sets and SAS programs. (2019-06-06)
Subject	Business and Management; Social Sciences

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>
    <relatedIdentifier relatedIdentifierType="DOI" relationType="HasPart">doi:10.7910/DVN/94EU5T/EZX4HN</relatedIdentifier>
    <relatedIdentifier relatedIdentifierType="DOI" relationType="HasPart">doi:10.7910/DVN/94EU5T/JCKHH3</relatedIdentifier>
```


Dataset and Files DOIs in metadata

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 Dataset Search

Updated Date | Download Format | Usage Rights | Free

H

Replication Data for: Using a Natural Experiment to Estimate...

dataverse.harvard.edu

Updated Sep 26, 2018

H

Data from: The UN Security Council Debates

dataverse.harvard.edu
search.datacite.org

Updated Jun 27, 2019

H

Data from: An optimistic outlook creates a rosy past: The impact...

dataverse.harvard.edu

Updated Dec 18, 2017

Data from: The UN Security Council Debates

Related Article

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

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

Unique identifier

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

Dataset updated Jun 27, 2019

Dataset provided by

Harvard Dataverse

License

CC0 1.0 Universal Public Domain Dedication

License information was derived automatically

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."

Jacobsen et al. 2016. *FAIR Principles: Interpretations and Implementation Considerations(Forthcoming)*

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 (e.g., access through cURL)
- Authentication API Tokens
- Authorization service

Getting Started with APIs

If you are a researcher or curator who wants to automate parts of your workflow, this section should help you get started. The Introduction section lists resources for other groups who may be interested in Dataverse APIs such as developers of integrations and support teams.

Contents:

Tokens and APIs described in Dataverse.org

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

Contact

Deaccession reason in dataset landing page when data not longer available

2000 Utah Colleges Exit Poll

Deaccessioned

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

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

TO BE INTEROBERABLE

"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."

Jacobsen et al. 2016. *FAIR Principles: Interpretations and Implementation Considerations(Forthcoming)*

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
- DDI (XML) as a rich schema to support extensive variable metadata

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 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, Harvard Dataverse, V1, UNF:6:CBIOoHJrG5/T6i+XjwBVwg== [fileUNF]

Cite Dataset

Learn about Data Citation Standards.

Variables 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

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 and 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

Extensive variable metadata automatically derived from tabular data file

Variable Price: Housing-c

Values	Categories	N

Summary Statistics

Cases	N
	2057
	0
Maximum	1.4085395336151123
Minimum	-0.7985830307006836
	-0.005094960725188625
	-0.06162228807806969
	0.357 3808803276065

Variable msaname:

Abilene,... 1

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

Text *

Date

YYYY-MM-DD

Subject *

- Agricultural Sciences
- Arts and Humanities
- Astronomy and Astrophysics
- Business and Management
- Chemistry

FAIR Controlled Vocabularies can be entered in metadata template

Keyword

Term

Vocabulary

Vocabulary URL

Enter full URL, starting with http://

Topic Classification

Term

Vocabulary

Vocabulary URL

Enter full URL, starting with http://

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 (*coming soon*)

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

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."

Jacobsen et al. 2016. *FAIR Principles: Interpretations and Implementation Considerations(Forthcoming)*

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

Replication Data for: Bean Counters: The Effect of Soy Tariffs on Change in Republican Vote Share between the 2016 and 2018 Elections

Version 1.0

Chyzh, Olga; R. Urbatsch, 2020, "Replication Data for: Bean Counters: The Effect of Soy Tariffs on Change in Republican Vote Share between the 2016 and 2018 Elections", https://doi.org/10.7910/DVN/CV7GYN, Harvard Dataverse, V1, UNF:6:ZhfyGSrlUw/89MqiIqKAmQ== [fileUNF]

Cite Dataset

Learn about Data Citation Standards.

Dataset Metrics

0 Downloads

Description

How do trade wars affect voting for the President's party? President Trump's aggressive tariffs on China, despite his largely rural electoral support base, provide a unique opportunity to analyze the relationship between international trade policy and domestic support. If trade-related considerations were ever decisive to American voters, the stark decrease in soy prices, a direct effect of Trump-initiated tariffs immediately preceding the 2018 midterm election, serves as a critical test for studying their effect. This letter shows a robust inverse relationship between county-level soybean production and the change in Republican vote share between the 2016 and 2018 congressional elections. (2019-09-05)

Subject Social Sciences

Keyword trade wars, tariffs, China, vote share, agriculture

CC0 as the default waiver for distributing open data

Files | Metadata | Terms | Versions

Terms of Use

Waiver

Our Community Norms as well as good scientific practices expect that proper credit is given via citation. Please use the data citation above, generated by the Dataverse.

CC0 - "Public Domain Dedication" PUBLIC DOMAIN

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

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

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)

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

Search this dataset... Find

Filter by

File Type: All Access: All

Sort

Metadata extracted from Astronomy FITS files

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

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		
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		

DATAVERSE FAIR SUMMARY

- Strong support for Findable, Accessible, and Reusable principles
- Weak for Interoperable principles
- In agreement* with FAIR test results (*F3 was fixed after test)
- There is no FAIR "compliance"
- Instead, it's a process and can always be improved

Beyond FAIR

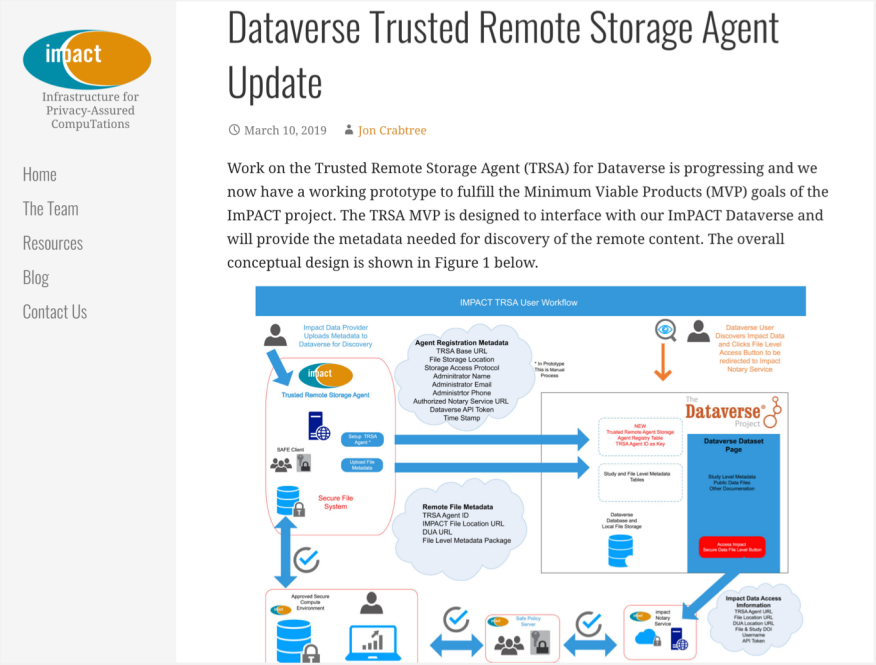
BEYOND FAIR

FAIR principles do not address:

- Responsible FAIR for sensitive data
- Data curation for data quality
- Capsules for reproducibility

Responsible FAIR: Current Efforts

TRUSTED REMOTE STORAGES



impact

Infrastructure for Privacy-Assured CompuTations

Home
The Team
Resources
Blog
Contact Us

Dataverse Trusted Remote Storage Agent Update

March 10, 2019 Jon Crabtree

Work on the Trusted Remote Storage Agent (TRSA) for Dataverse is progressing and we now have a working prototype to fulfill the Minimum Viable Products (MVP) goals of the ImPACT project. The TRSA MVP is designed to interface with our ImPACT Dataverse and will provide the metadata needed for discovery of the remote content. The overall conceptual design is shown in Figure 1 below.

http://cyberimpact.us/

- Trusted Remote Storage Agents
- Notary Service for auth/authz
- Dataverse stores only metadata
- IMPACT project funded by NSF

STANDARD DATA SHARING AND USE POLICIES

DataTags:

Blue: Open, no authentication needed

Green: Authentication; No authorization needed

Yellow: Authentication and Authorization; Data might be downloadable

Orange: Institution DUA needed; Data might be in TRSA

Red: Only Metadata in Dataverse; Data not downloadable

Crimson: Only Metadata in Dataverse; Data outside network

STANDARD DATA SHARING AND USE POLICIES

March 5th Workshop on “Standardizing Data Sharing, Use, and Access Agreements” hosted by Microsoft/Harvard Dataverse (by invitation only)

PRIVACY PRESERVING TOOLS: OPENDP

https://privacytools.seas.harvard.edu/

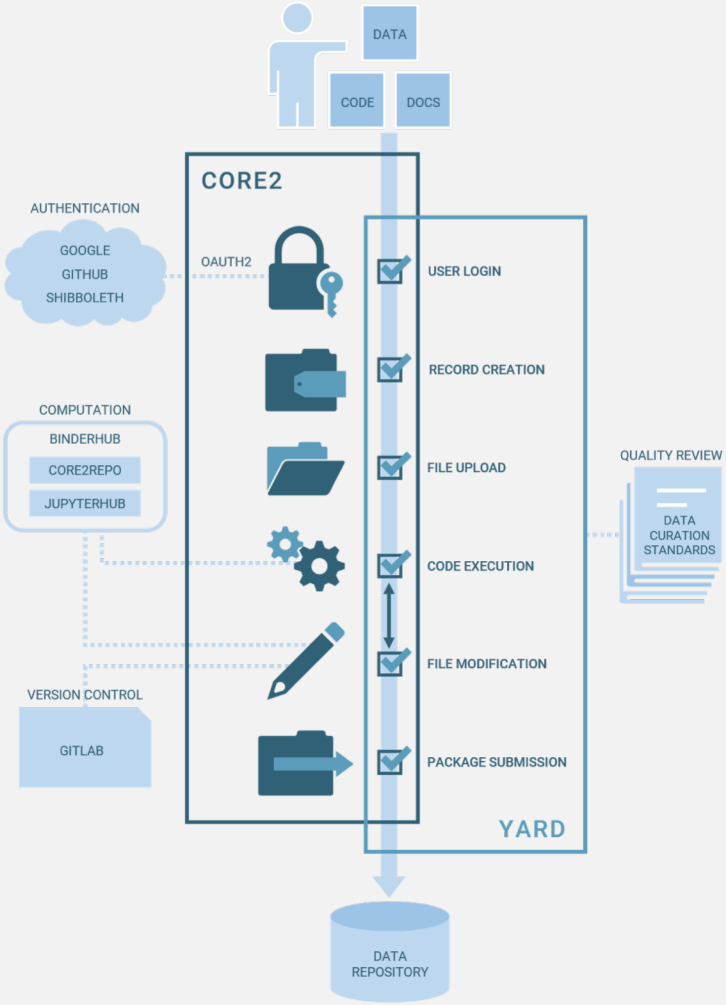
- Trustworthy differential privacy tools suite
- A community, open-source effort
- For statistical research on sensitive data
- To be launched in 2020 with Sloan funding
- A use case: sensitive data in Dataverse

Initially led by Harvard Privacy Tools

(PIs: Vadhan, King, Honaker, Crosas)

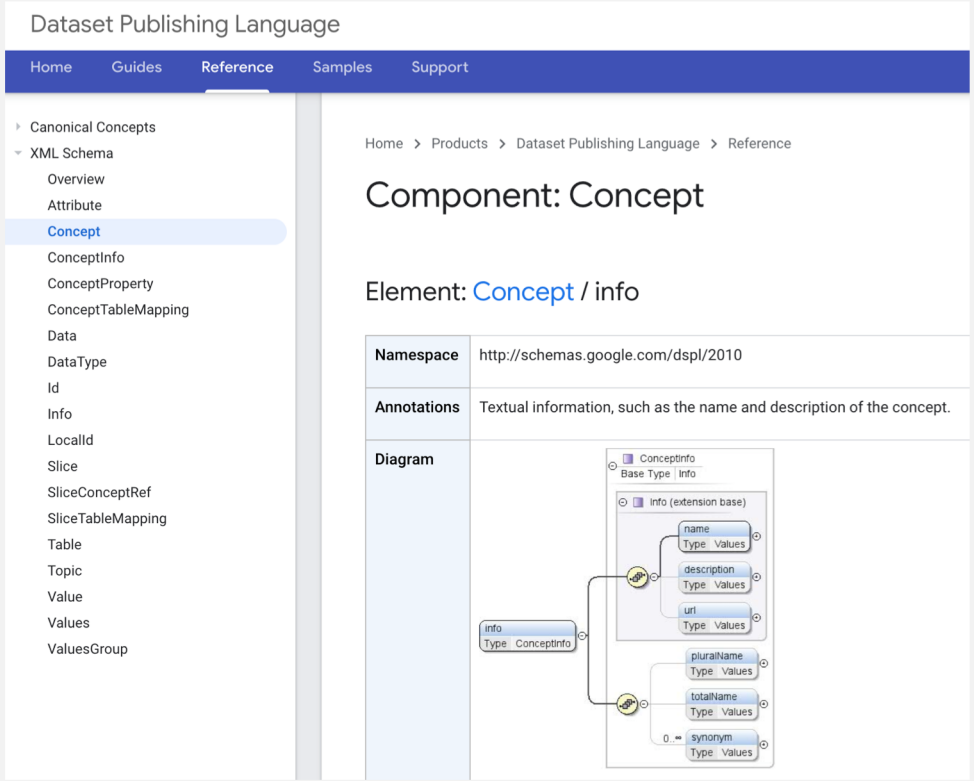
Data Quality: Current Efforts

DATA CURATION TOOLS



- Combine CORE2 and YARD
- Curation and verification workflow
- Quality review based on standards
- PIs: Peer, Christian, Crabtree, Crosas

METADATA AND VOCABULARIES STANDARDS

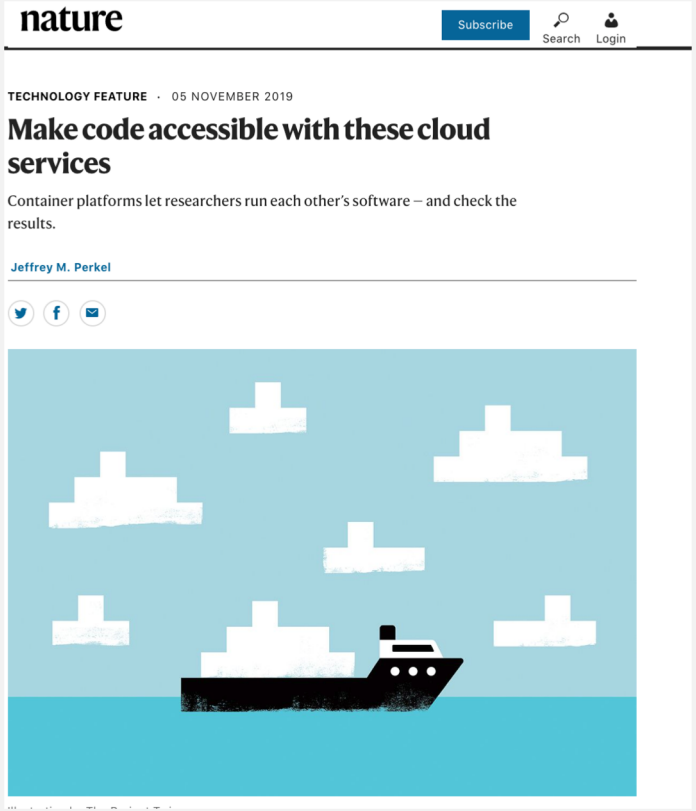


- Better standard metadata support
- Support data vocabularies
- Consider DDI update
- Consider Google's Dataset Publishing Language

https://developers.google.com/public-data

Reproducibility: Current Efforts

CONTAINER SUPPORT



nature

Subscribe Search Login

TECHNOLOGY FEATURE · 05 NOVEMBER 2019

Make code accessible with these cloud services

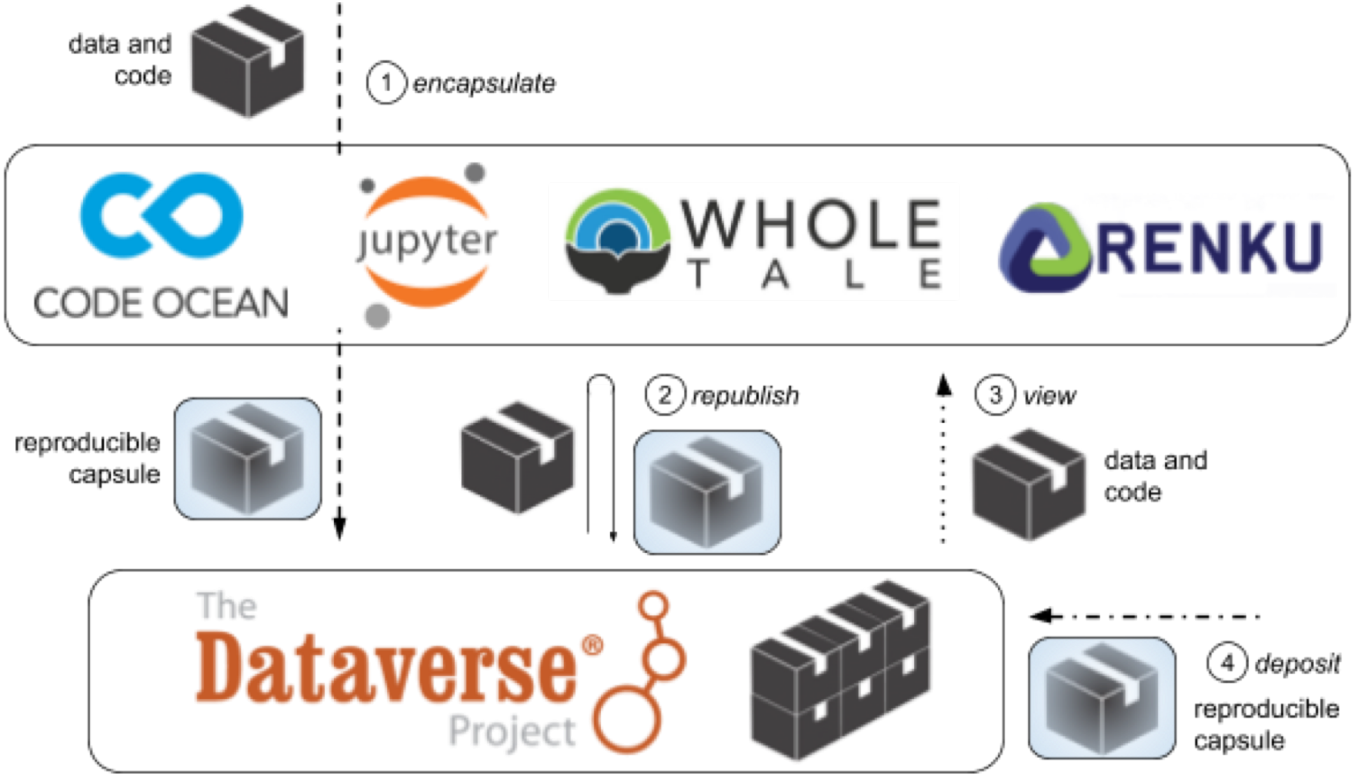
Container platforms let researchers run each other's software — and check the results.

Jeffrey M. Perkel

- Containers becoming a standard for computational reproducibility
- Add container support
- Integrate with container platforms

https://www.nature.com/articles/d41586-019-03366-x

CONTAINER SUPPORT USES CASES



Trisovic, Crosas, et al, 2020, working paper

SUMMARY

- **FAIR** enables machine-actionability of data resources
- Dataverse has currently strong support for **FAR**, but not **I**
- Data creators, stewards, curators need to do their part
- **Beyond FAIR**, Dataverse repositories plans to support:
- Responsible sharing of sensitive data
- Integration with curation tools to improve data quality
- Integration with reproducible tools to verify results

THANKS @mercecrosas