

The Dataverse **Project for** managing and sharing your research data

Morning Agenda, 02/20/2023

- 9:30am -9:45am
 - Introduction and brief high level information on today's agenda
 - Stefano, Gustavo, Sonia, Phil, Stephen
- 9:45am 10:10am
 - Dataverse Software New Features and Future Plans (Gustavo)
- 10:10am 10:30am
 - APIs and Integrations (Phil)
- **10:30-10:45 -** 15 minute break
- 10:45am-12:30pm (Sonia)
 - How The Dataverse software supports data sharing
 - High level basics of curation in a Dataverse based repository
 - Account creation, Dataverse collection creation and features, Dataset creation and features, and file upload and features

Afternoon Agenda, 02/20/2023

- 12:30pm -1:30pm Lunch Break
- 1:40pm 2:40pm (Sonia)
 - Quality Assessment of DRAFT datasets, draft sharing options, Publishing, versioning
- 2:40 3:00pm Break/Questions and Conversations
- 3-4:30pm
 - O Hands-on:
 - Create a collection
 - Deposit a dataset
 - Upload files
 - Review Draft
 - Publish
- **4:30-4:50pm** (Stephen)
 - Creating Bug Reports and Change Requests
- **4:50pm 5:15pm** Questions and conversations



What is the Purpose of a data repository?

The **purpose of a data repository** is to keep a certain population of data so that it can be mined for greater insight or business intelligence or to be used for a specific reporting need. https://en.wikipedia.org/wiki/Information_repository

The Registry of Research Data Repositories (re3data.org) is an Open Science tool that offers researchers, funding organizations, libraries and publishers an overview of existing international repositories for research data. https://en.wikipedia.org/wiki/Registry of Research Data Repositories

Data should be submitted to discipline-specific, community-recognized repositories where possible. Where a suitable discipline-specific resource does not exist, data should be submitted to a generalist repository.

Open source research data repository software

Establish a research data management solution for your community. Federate with a growing list of Dataverse repositories worldwide for increased discoverability of your community's data.

Researchers enjoy full control over your data. Receive web visibility, academic credit, and increased citation counts.

Journals seamlessly manage the submission, review, and publication of data associated with published articles

Developers contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization* and *exploration tools,* or other research and data archival systems with the Dataverse Project.



IQSS leads the development of the open source Dataverse Project software

The <u>Open Data Assistance Program at Harvard</u> (a collaboration with **Harvard Library**, the **Office for Scholarly Communication** and IQSS), provides user support.

The Library Technology Services at HUIT provides hosting and backups support of the Harvard Dataverse Repository.



Community

- CommunityMeetings
- Community Calls
- Global Dataverse
 Community
 Consortium

Best Practices

- Academic Credit
- Data Citation
- <u>Dataverse</u><u>Community Norms</u>
- Data Management
- Replication Dataset
 Guidelines

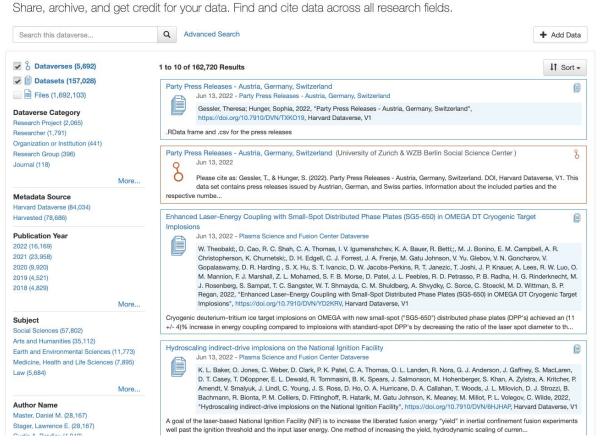
Software

- Goals, Roadmap, and Releases
- Collaborations
- <u>Integrations</u>
- Features
- Source Code
- Guides



38.666.339 Downloads

Metrics



User Guide

Log In

Contact C Share

Log In

帮助

首页

数据资源

新闻动态

分类统计

国际资源

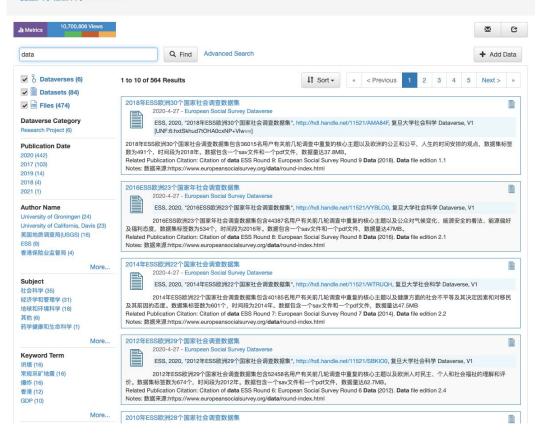
代言系列

平台简介

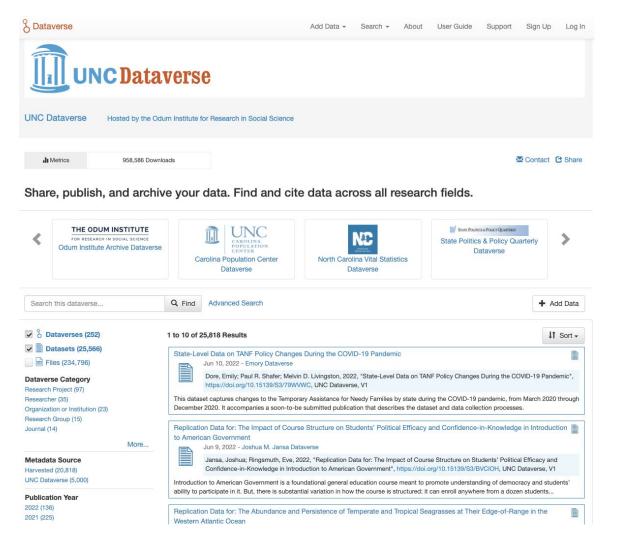
相关链接

联系我们

复旦大学社会科学 Dataverse











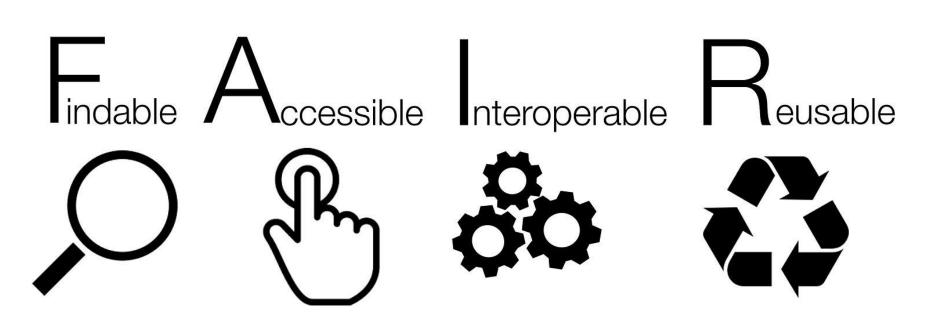
Today's Agenda

What is The Dataverse Project?

How the Dataverse software supports research data sharing

Dataverse collections, datasets, and files

Dataverse and FAIR data principles



https://en.wikipedia.org/wiki/File:FAIR_data_principles.jpg



- Support for FAIR Data Principles
- Data citation for datasets and files
- OAI-PMH (Harvesting)
- APIs for interoperability and custom integrations
- API client libraries
- DataCite integration
- Login via Shibboleth
- Login via ORCID, Google, GitHub, or Microsoft
- Login via OpenID Connect (OIDC)
- Internationalization
- Versioning
- Restricted files
- Embargo
- Custom licenses
- Custom terms of use
- Publishing workflow support
- File hierarchy

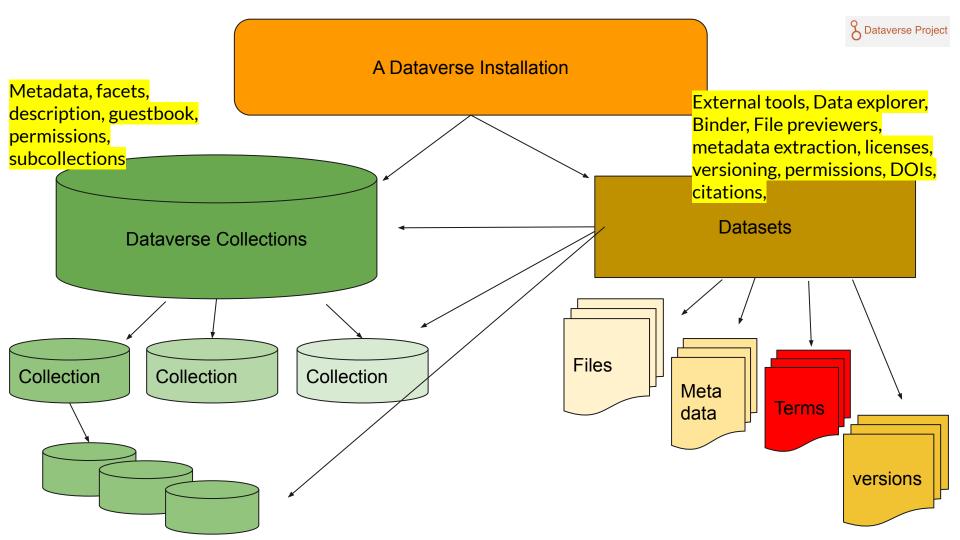
- File previews
- Preview and analysis of tabular files
- Usage statistics and metrics
- Guestbook
- Fixity checks for files
- File download in R and TSV format
- Faceted search
- Customization of collections
- Private URL
- Widgets
- Notifications
- Schema.org JSON-LD
- External tools
- External vocabulary
- Dropbox integration
- GitHub integration
- Integration with Jupyter notebooks

- User management
- Curation status labels
- **Branding**
- Backend storage on S3 or Swift
- Direct upload and download for \$3

Export data in BagIt format

- Post-publication automation (workflows)
- Pull header metadata from Astronomy (FITS) files
- Provenance
- Support for rsync
- Auxiliary files for data files

https://dataverse.org/software-features





Users/Groups ❤ All the users and groups that have access to your dataverse.
Roles 📤 All the roles set up in your dataverse, that you can assign to users and groups.
+ Add New Role
Admin - A person who has all permissions for dataverses, datasets, and files. AddDataverse AddDataset ViewUnpublishedDataverse ViewUnpublishedDataset DownloadFile EditDataverse EditDataset ManageDataversePermissions ManageDatasetPermissions PublishDataverse PublishDataset DeleteDataverse Dele
Contributor - For datasets, a person who can edit License + Terms, and then submit them for review. ViewUnpublishedDataset
Curator - For datasets, a person who can edit License + Terms, edit Permissions, and publish datasets. AddDataset ViewUnpublishedDataset ViewUnpublishedDataset ViewUnpublishedDataset ViewUnpublishedDataset DownloadFile EditDataset ManageDatasetPermissions PublishDataset DeleteDatasetDraft
Dataset Creator - A person who can add datasets within a dataverse. AddDataset
Dataverse + Dataset Creator - A person who can add subdataverses and datasets within a dataverse. AddDataverse AddDataset
Dataverse Creator - A person who can add subdataverses within a dataverse. AddDataverse
File Downloader - A person who can download a published file. DownloadFile
Member - A person who can view both unpublished dataverses and datasets. ViewUnpublishedDataverse ViewUnpublishedDataset DownloadFile

Edit Access *

Who can add to this dataverse?

- Anyone adding to this dataverse needs to be given access
- Anyone with a Dataverse account can add sub dataverses
- Anyone with a Dataverse account can add datasets
- Anyone with a Dataverse account can add sub dataverses and datasets

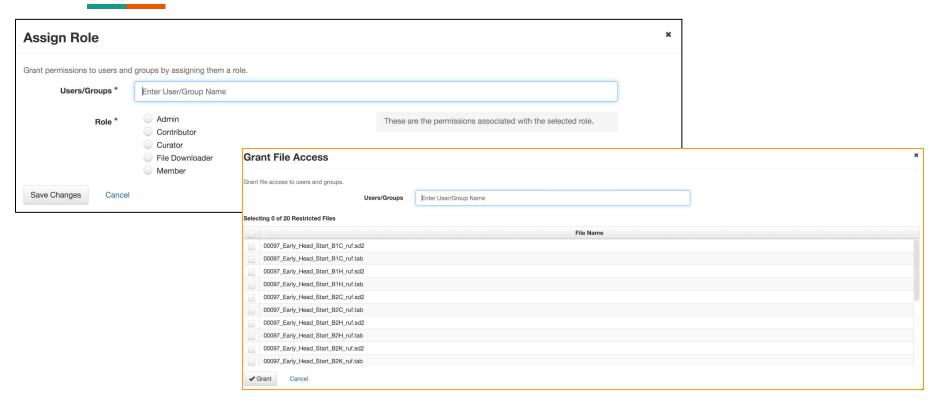
When a user adds a new dataset to this dataverse, which role should be automatically assigned to them on that dataset?

- Ontributor Edit metadata, upload files, and edit files, edit Terms, Guestbook, Submit datasets for review
- Curator Edit metadata, upload files, and edit files, edit Terms, Guestbook, File Restrictions (Files Access + Use), Edit Permissions/Assign Roles + Publish

Save Changes

Cancel







Today's Agenda

What is The Dataverse Project?

How the Dataverse software supports research data sharing

Dataverse collections, datasets, and files

Dataverse and FAIR data principles

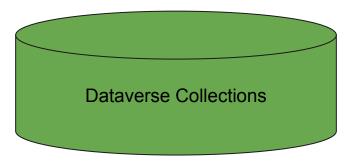


Dataverse Collections

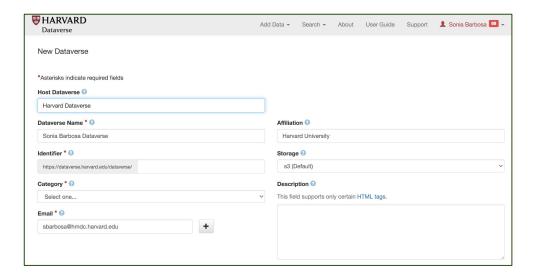
- General Information edit your metadata, facets, description
- Theme + Widgets customize, add widget to your website
- Permissions
- Groups
- Dataset Templates customize consistent data deposits, terms, etc...
- Dataset Guestbooks who is using your data and why?
- Featured Dataverses highlight your nested Dataverse collections



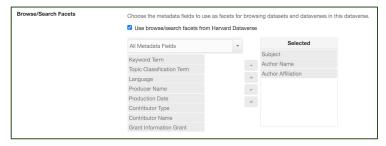
- Metadata
- Searchable facets
- Holds datasets

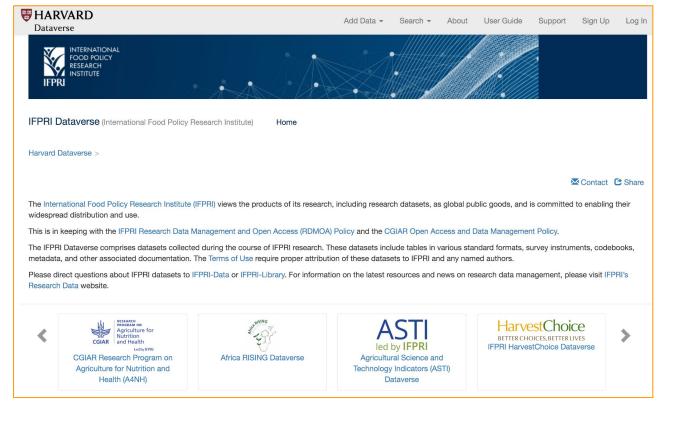


Customizable with descriptions, logos, collection space URLs, affiliation, contact emails, permissions, metadata and searchable facets. **Dataverse collections hold datasets.**













Files (11,208) **Dataverse Category**

Dataverses (9)

Datasets (602)

Research Project (8)

Research Group (1) **Publication Year**

2021 (36) 2020 (97) 2019 (75) 2018 (66)

2017 (100)

More...

Topic Classification Term

Agricultural research ((AGROVOC)) (120)

Nutrition ((AGROVOC)) (104) Impact assessment ((AGROVOC)) (88)

Health ((AGROVOC)) (69)

Social accounting matrix ((STW Thesaurus for

Economics)) (65)

More...

Keyword Term

AFRICA (215) AFRICA SOUTH OF SAHARA (203)

EAST AFRICA (114) nutrition (98)

Africa (96)

More...

Series Name

Household- and Community-level Surveys

(262)

Country Level Data (88)

Social Accounting Matrix (SAM) (69)

Biophysical Surveys (41) Experimental Data (34)

More...

Geographic Coverage Country / Nation

Ethiopia (77)

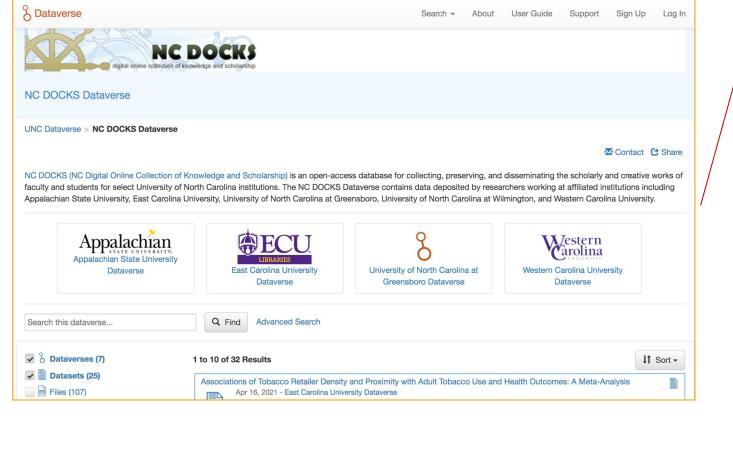
Tanzania, United Republic of (62)

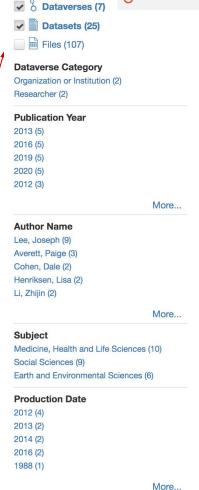
Bangladesh (60)

Ghana (46) Malawi (43)

More...

Grant Information Grant Agency

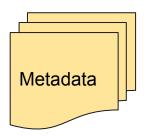




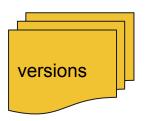
2 Dataverse Project

Datasets









Datasets - citations





Harvard Dataverse > IFPRI Dataverse > Africa RISING Dataverse >

Assessment of Land Degradation in Semi-Arid Zone of Central Tanzania

Version 1.0



International Institute of Tropical Agriculture (IITA); University of Bonn, Germany, 2021, "Assessment of Land Degradation in Semi-Arid Zone of Central Tanzania", https://doi.org/10.7910/DVN/HADLCK, Harvard Dataverse, V1, UNF:6:jU30uW/R5JgiHw7dPr5iuQ== [fileUNF]

Cite Dataset -

Learn about Data Citation Standards.

Data Citation

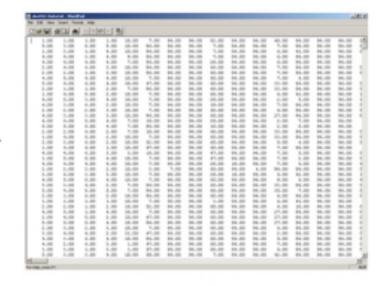
Your Publication



Formal Data Citation



Your Data





Principle 2
- Credit and
Attribution:
Such as authors,
repositories or
other distributors
and contributors.

Author(s), Year, Dataset Title,
 Global Persistent Identifier, —
 Data Repository or Archive,
 Version

Principle 7 - Specificity and verification: Such as the specific version used. Versioning or timeslice information should be supplied with any updated or dynamic dataset. Principles 4, 5, 6
- Unique Identification,
Access, Persistence:
A unique, persistent
identifier, such as a DOI
or Handle, that provides
access to metadata.

Data Citation

A data citation in a Dataverse repository has seven components:

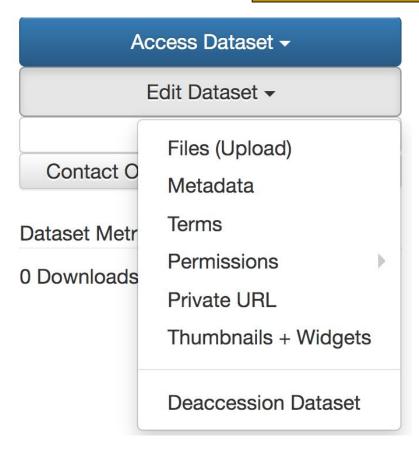
- author name(s)
- year (date published in the Dataverse repository)
- title
- global persistent identifier: <u>DOI</u> or <u>Handle</u>
- publisher (repository that published the dataset)
- version number
- <u>universal numerical fingerprint (UNF)</u>: for tabular data (being phased out)

Here is an example of a replication data citation from Political Analysis, Hanmer, Banks, and White, 2013:

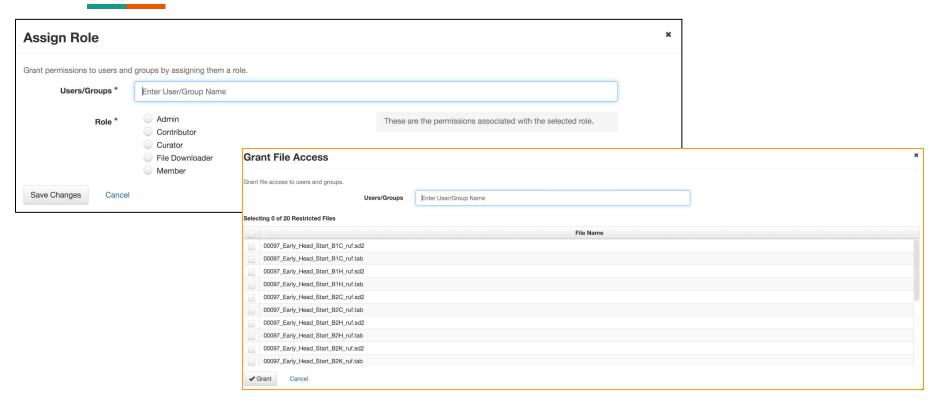
Hanmer, Michael J.; Banks, Antoine J., White, Ismail K., 2013, "Replication data for: Experiments to Reduce the Over-reporting of Voting: A Pipeline to the Truth", doi:10.7910/DVN/22893, Harvard Dataverse, V2, UNF:5:eJOVAjDU0E0jzSQ2bRCg9g==

See more detailed information about the Dataverse Project's <u>Data Citation Standard</u>.

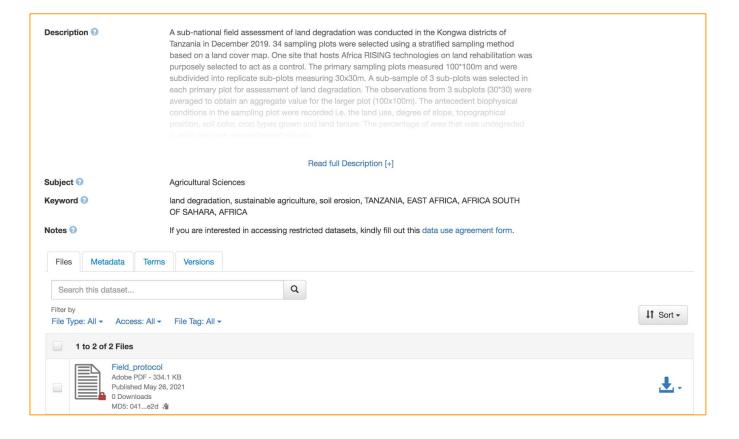
Datasets





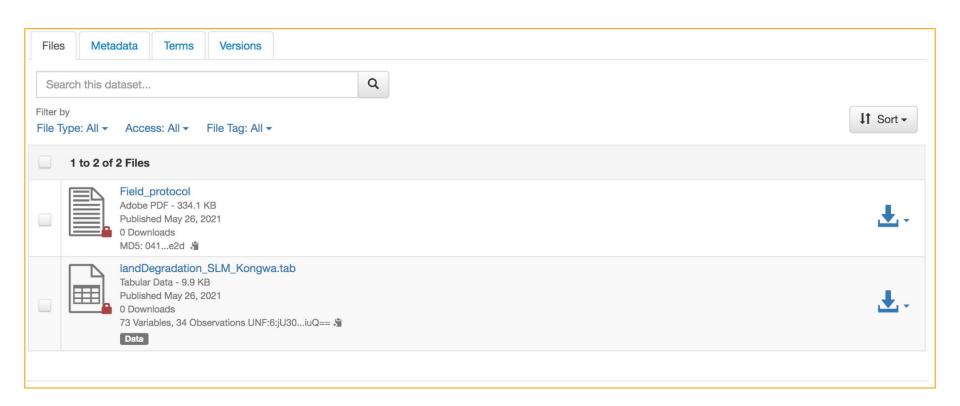


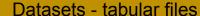
Datasets



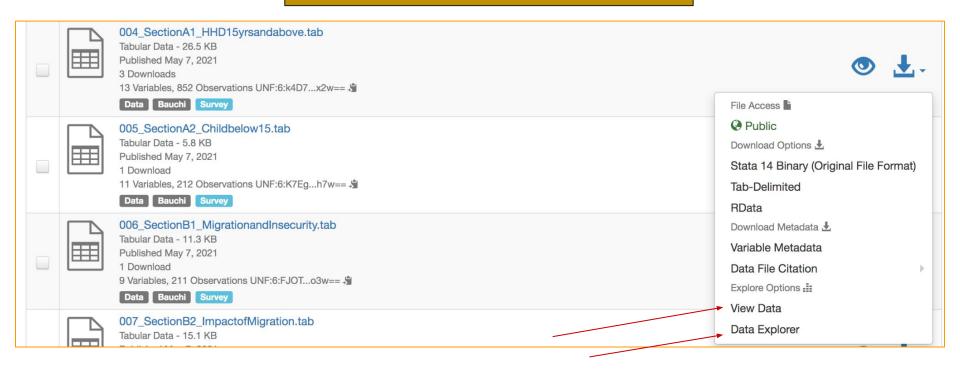
Datasets - files











Datasets - tabular files

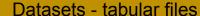


Supported File Formats

Tabular Data ingest supports the following file formats:

File format	Versions supported					
SPSS (POR and SAV formats)	7 to 22					
STATA	4 to 15					
R	up to 3					
Excel	XLSX only (XLS is NOT supported)					
CSV (comma-separated values)	(limited support)					

See the subsections in the left sidebar for more information on each of these supported formats.





Third National Fadama Development Financing II Impact Study Household Survey in Bauchi 004 SectionA1 HHD15vrsandabove.tab



/ariable educiv	l: Highest Level of e	ducation				
Values	Categories					
8	Koranic education	on .				
5	Some secondary	reducation (Incl. Junior secondary school)				
2	Adult literacy tra					
1	No formal educa	tion				
99	Others					
7	Post-secondary	ndary education				
6	Completed seco	ndary education				
3	Some primary e	ducation				
4	Completed prim	ary education				
	ctvty16: Other prima	ary activity(specify)in 2016 Categorie				
Values Summary Statis	tics	Categorie				
Values Summary Statis Variable priacts	tics ty16: Primary activit	Categorie y in 2016				
Summary Statis /ariable priacts	tics ty16: Primary activit	Categorie y in 2016 tegories				
Summary Statis /ariable priact Values	tty16: Primary activit Ca	Categorie y in 2016 tegoriee dent in school (any type)				
Summary Statis /ariable priacty Values 14	ty16: Primary activit Ca Sh	Categorie y in 2016 tegories dent in school (any type) estock production				
Summary Statis /ariable priacts Values 14 2 7	tics ty16: Primary activit Ca St. Liv. Tra	V in 2016 tegories dent in school (any type) estock production resportation business				
Summary Statis fariable priacts Values 14 2 7	titics ty16: Primary activit Ca Sit Liv Tra Co	V in 2016 tegories dent in school (any type) setock production resportation business				
Summary Statis fariable priacts Values 14 2 7 10 18	tics ty18: Primary activit Ca St Liv Tra Co Co	Categorie y in 2016 tegoriee dent in school (any type) setook production nesportation business nestruction tsians(incl. Mechanics)				
Values Summary Statis /ariable priacts Values 14 2 7 10 18	tios ty16: Primary activit Ca St. Liv. Tra Co Co Art	V in 2016 tegories dent in school (any type) setook production nsportation business nstruction sians(Incl. Mechanics)				
Values Summary Statis	ty16: Primary activit Ca St. Liv Tra Co Art File	Categorie y in 2016 tegoriee dent in school (any type) setook production nesportation business nestruction tsians(incl. Mechanics)				
Values Summary Statis Variable priacts 14 2 7 10 18 3 12	tics ty16: Primary activit Ca Sta Liv Tra Co Art Fis	y in 2016 tegories dent in school (any type) estock production nsportation business nstruction sisinal[inol. Mechanics)				

https://guides.dataverse.org/en/latest/user/tabulardataingest/index.html https://github.com/scholarsportal/dataverse-data-explorer-v2



Datasets - file previewer

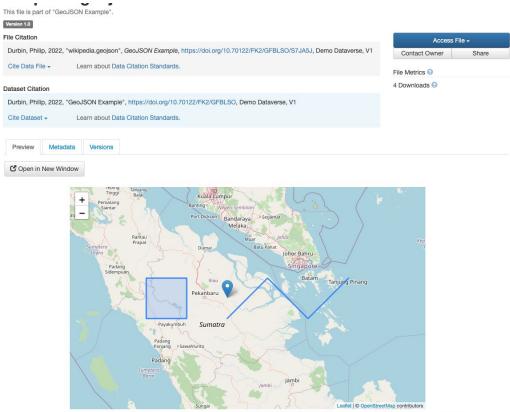
Preview Metadata Versions

Explore on View Data

	householdid	memname	age	reltohhh	otherreltohh	gender	educlvl	othereducivi	priactvty16	othpriactvty16	priactvty18
5	2060154		25	3		1	7		14		14
6	2060154		18	3		2	5		14		14
7	2060156		61	1		1	4		6		1
8	2060156		17	3		2	6		14		13
9	2060156		15	3		1	6		14		16
10	2060156		32	2		2	8		18		18
11	2060157		52	1		1	4		5		5
12	2060157		37	2		2	4		5		5
13	2060157		32	3		1	6		5		5
14	2060157		32	3		2	4		8		8
15	2060158		48	1		1	6		5		5
16	2060158		25	2		2	4		8		8
17	2060158		17	3		1	5		14		14
18	2060158		16	3		1	5		5		14
19	2060159		48	1		2	4		7		5
20	2060159		21	3		1	5		6		14
21	2060159		17	3		2	5		6		14
22	2060159		15	3		2	4		2		13
23	2060160		80	1		1	3		3		4
24	2060160		50	2		2	4		6		4
25	2060160		30	3		1	5		8		16
26	2060160		25	3		2	4		5		16
27	2060161		70	1		1	4		8		8
28	2060161		50	2		2	1		8		5
29	2060161		30	3		1	6		5		8
30	2060152		27	3		1	6		5		14
31	2060153		48	1		2	4		1		8



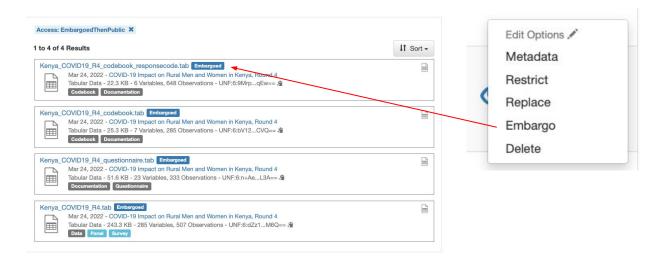
Datasets - GeoJSON previewer



Previewers originally developed by QDR and maintained at https://github.com/GlobalDataverseCommunityConsortium/dataverse-previewers. Feedback and contributions welcome.



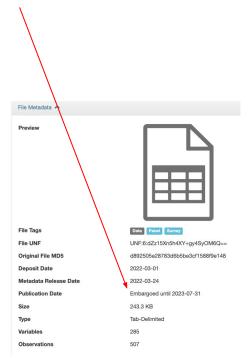
Datasets - file level embargo



Harvard Dataverse > IFPRI Dataverse >

COVID-19 Impact on Rural Men and Women in Kenya, Round 4





https://github.com/GlobalDataverseCommunityConsortium/dataverse-previewers

Datasets - metadata



Detailed below are what metadata schemas we support for Citation and Domain Specific Metadata in the Dataverse Project:

- Citation Metadata: compliant with DDI Lite, DDI 2.5 Codebook, DataCite 3.1, and Dublin Core's DCMI Metadata Terms (see .tsv version). Language field uses ISO 639-1 controlled vocabulary.
- Geospatial Metadata: compliant with DDI Lite, DDI 2.5 Codebook, DataCite, and Dublin Core (see .tsv version). Country / Nation field uses ISO 3166-1 controlled vocabulary.
- Social Science & Humanities Metadata: compliant with DDI Lite, DDI 2.5 Codebook, and Dublin Core (see .tsv version).
- Astronomy and Astrophysics Metadata: These metadata elements can be mapped/exported to the International Virtual
 Observatory Alliance's (IVOA) VOResource Schema format and is based on Virtual Observatory (VO) Discovery and Provenance
 Metadata (see .tsv version).
- Life Sciences Metadata: based on ISA-Tab Specification, along with controlled vocabulary from subsets of the OBI Ontology and the NCBI Taxonomy for Organisms (see .tsv version).
- Journal Metadata: based on the Journal Archiving and Interchange Tag Set, version 1.2 (see .tsv version).

See also the Dataverse Software 4.0 Metadata Crosswalk: DDI, DataCite, DC, DCTerms, VO, ISA-Tab document and the Metadata Customization section of the Admin Guide.

Datasets - metadata

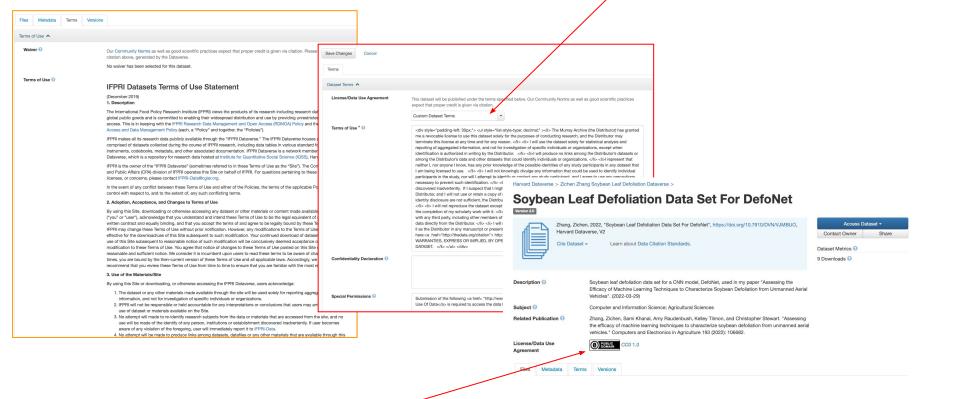


Metadata Fields	Choose the metadata fields to use in dataset templates and when adding a dataset to this dataverse.
	Use metadata fields from Harvard Dataverse
	Citation Metadata (Required) [+] View fields
	☐ Geospatial Metadata [+] View fields
	☐ Social Science and Humanities Metadata [+] View fields
	☐ Astronomy and Astrophysics Metadata [+] View fields
	☐ Life Sciences Metadata [+] View fields
	☐ Journal Metadata [+] View fields

		Export Metadata •				
Citation Metadata ^		Dublin Core DDI				
Dataset Persistent ID 🕣	doi:10.7910/DVN/HADLCK	DataCite DDI HTML Codebook				
Publication Date	2021-05-26	JSON				
Title 🕣	Assessment of Land Degradation in Semi-Arid Zone of Central Tanzania	OAI_ORE				
Author ③	International Institute of Tropical Agriculture (IITA) University of Bonn, Germany	OpenAIRE Schema.org JSON-LD				
Contact	Use email button above to contact.					
	Muthoni, Francis (International Institute of Tropical Agriculture (IITA))					
Description	A sub-national field assessment of land degradation was conducted in the Kongwa districts of Tanzania in December 2019. 34 sampling plots were selected using a stratified sampling method based on a land cover map. One site that hosts Africa RISING technologies on land rehabilitation was purposely selected to act as a control. The primary sampling plots measured 100*100m and were subdivided into replicate sub-plots measuring 30x30m. A sub-sample of 3 sub-plots was selected in each primary plot for assessment of land degradation. The observations from 3 subplots (30°30) were averaged to obtain an aggregate value for the larger plot (100x100m). The antecedent biophysical conditions in the sampling plot were recorded i.e. the land use, degree of slope, topographical position, soil color, crop types grown and land tenure. The percentage of area that was undegraded in each plot was also estimated visually.					
	A questionnaire for mapping land degradation and sustainable land management was applied for v extent, degree, and direct causes of land degradation. The different types of land degradation, for cscord whether they are present or not, in addition to their extent and degree (intensity). The extent sub-plot covered by different types of land degradation, the degree was divided into 4 classes in as land degradation (0 = Light, 1 = Moderate, 2 = Strong, 3 = Severe). Moreover, the type, purpose, ex sustainable land management practices (SLM) were visually assessed in every subplot. Data were KoboCollect toolbox and transmitted to a cloud database for storage and descriptive analysis.	example erosion by water, were represented the proportion of a scending order of intensity of ttent, and effectiveness of				
Subject ()	Agricultural Sciences					
Keyword ©	land degradation (AGROVOC) http://aims.fao.org/aos/agrovoc/c_34823 sustainable agriculture (AGROVOC) http://aims.fao.org/aos/agrovoc/c_33561 soil erosion (AGROVOC) http://aims.fao.org/aos/agrovoc/c_2651 TANZANIA (AGROVOC) http://aims.fao.org/aos/agrovoc/c_7608 EAST AFRICA (AGROVOC) http://aims.fao.org/aos/agrovoc/c_2442 AFRICA SOUTH OF SAHARA (AGROVOC) http://aims.fao.org/aos/agrovoc/c_166 AFRICA (AGROVOC) http://aims.fao.org/aos/agrovoc/c_165					
Topic Classification ()	Natural resources (AGROVOC) http://aims.fao.org/aos/agrovoc/c_5091 Sustainable agriculture (AGROVOC) http://aims.fao.org/aos/agrovoc/c_33561					
Notes ②	If you are interested in accessing restricted datasets, kindly fill out this data use agreement form.					
Language 0	English					
Producer	International Institute of Tropical Agriculture (IITA) https://www.iita.org/					



Datasets - terms



Datasets - versions



Datasets - deaccession

Replication Data for: Democratic Subversion: Elite Cooptation and Opposition Fragmentation Deaccessioned





Arriola, Leonardo R.; DeVaro, Jed; Meng, Anne, 2021, "Replication Data for: Democratic Subversion: Elite Cooptation and Opposition Fragmentation", https://doi.org/10.7910/DVN/GHDKSY, Harvard Dataverse, V1, DEACCESSIONED VERSION, UNF:6:Nk/xhGPVreurGI55KFcKrQ== [fileUNF]

The dataset has been transferred to another repository



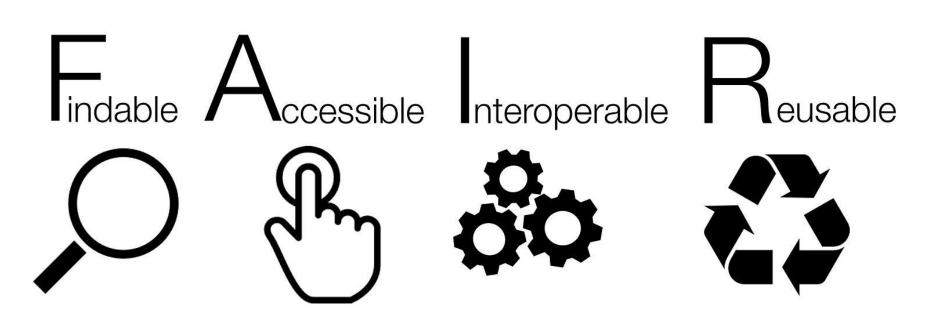
Today's Agenda

What is The Dataverse Project?

How the Dataverse software supports research data sharing

Dataverse collections, datasets, and files

Dataverse and FAIR data principles



https://en.wikipedia.org/wiki/File:FAIR_data_principles.jpg



Thank you! support@dataverse.org

The Dataverse Project https://dataverse.org/

The Harvard Dataverse repository https://dataverse.harvard.edu/

Integrating Dataverse and DSpace https://osf.io/72w4m

The Dataverse Guide https://guides.dataverse.org/en/latest/

Dataverse on Github https://github.com/IQSS/dataverse

Dataverse google community https://groups.google.com/g/dataverse-community