# The Dataverse Project for managing your research data

Sonia Barbosa
Manager of Data Curation, Harvard Dataverse
Manager of the Murray Archive

Dataverse Project

# Today’s Agenda

## What is The Dataverse Project?

The Harvard Dataverse Repository

How the Dataverse software supports research data sharing

Dataverse collections, datasets, and files

Dataverse and FAIR data principles

# The Dataverse Project

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.

# The Dataverse Project

**IQSS leads the development** of the open source Dataverse Project software

The Open Data Assistance Program at Harvard (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.

# The Dataverse Project

## Community

- Community Meetings
- Community Calls
- **Global Dataverse Community Consortium**

## Best Practices

- Academic Credit
- Data Citation
- Dataverse Community Norms
- Data Management
- Replication Dataset Guidelines

## Software

- Goals, Roadmap, and Releases
- Collaborations
- Integrations
- Features
- Source Code
- Guides

DATAVERSE REPOSITORIES - A WORLD VIEW

69 Installations

# Today’s Agenda

**What is The Dataverse Project?**

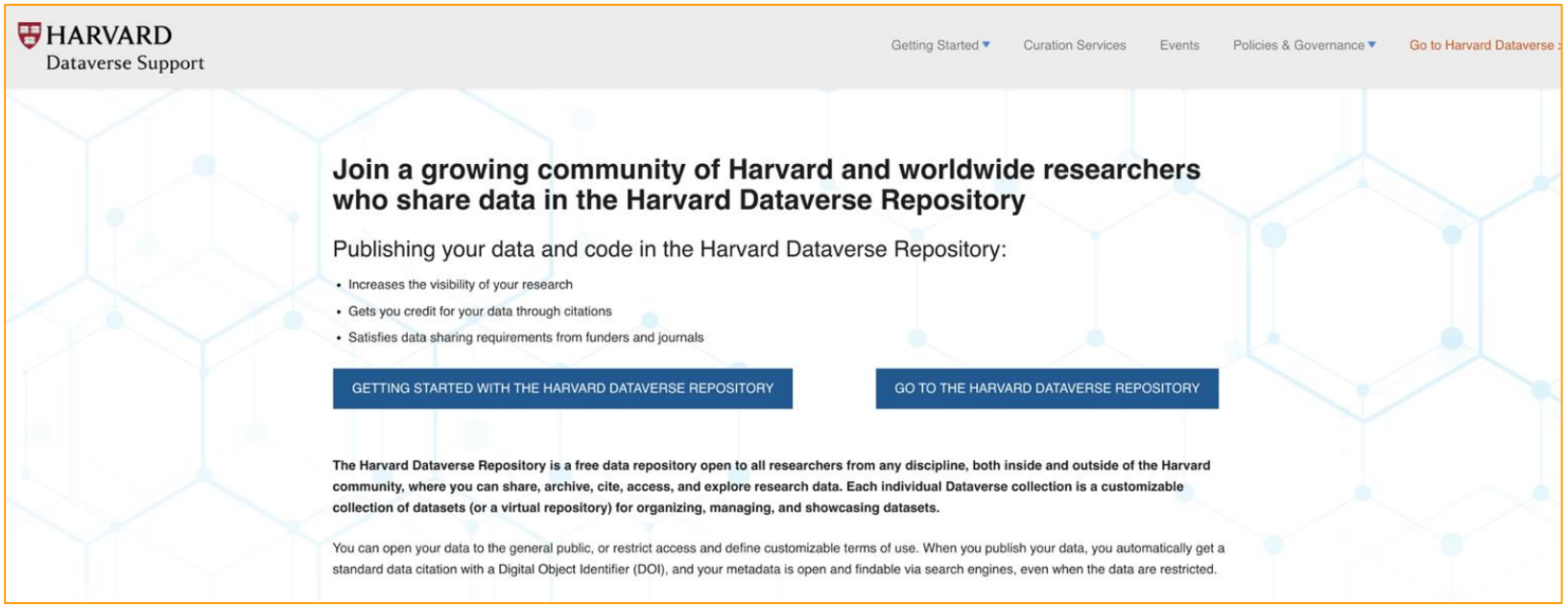
## The Harvard Dataverse Repository

How the Dataverse software supports research data sharing

Dataverse collections, datasets, and files

Dataverse and FAIR data principles

# The Harvard Dataverse Repository




dataverse.harvard.edu

# Today’s Agenda

**What is The Dataverse Project?**

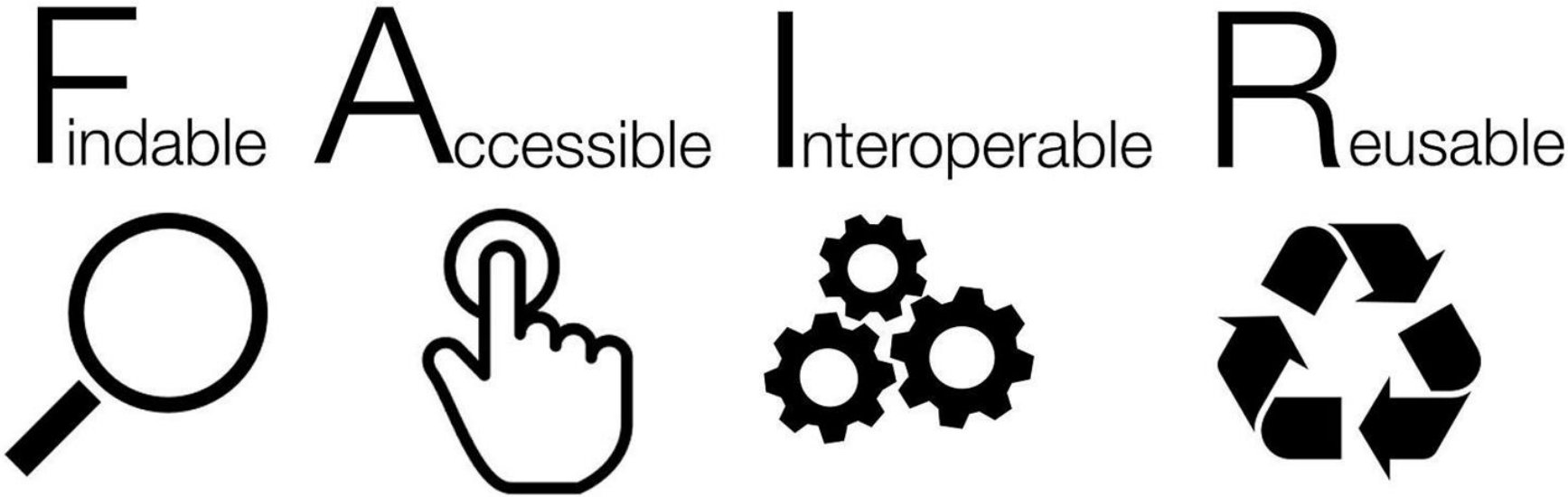
**The Harvard Dataverse Repository**

## How the Dataverse software supports research data sharing

Dataverse collections, datasets, and files

Dataverse and FAIR data principles

# FAIR principles and Dataverse

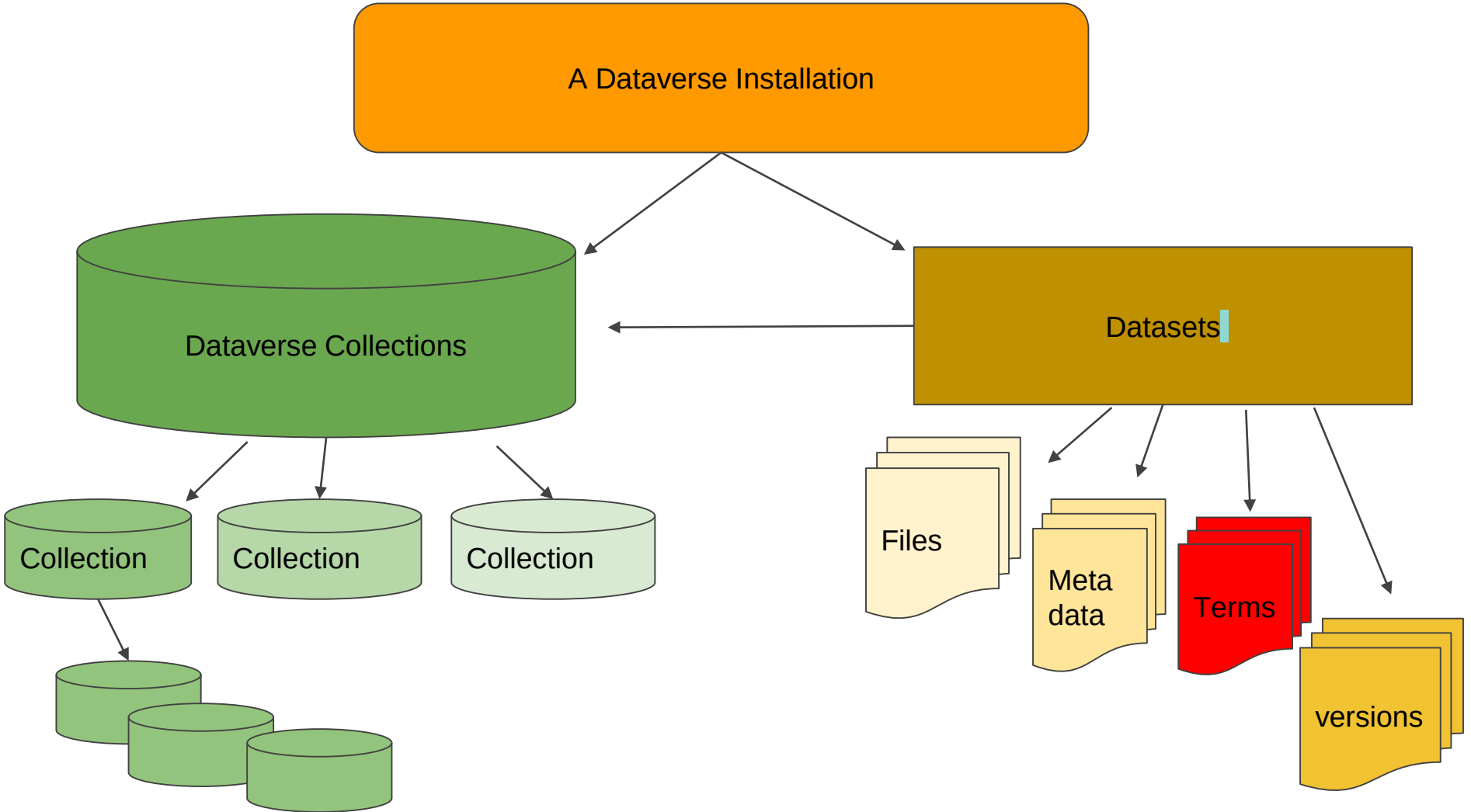



https://en.wikipedia.org/wiki/File:FAIR_data_principles.jpg

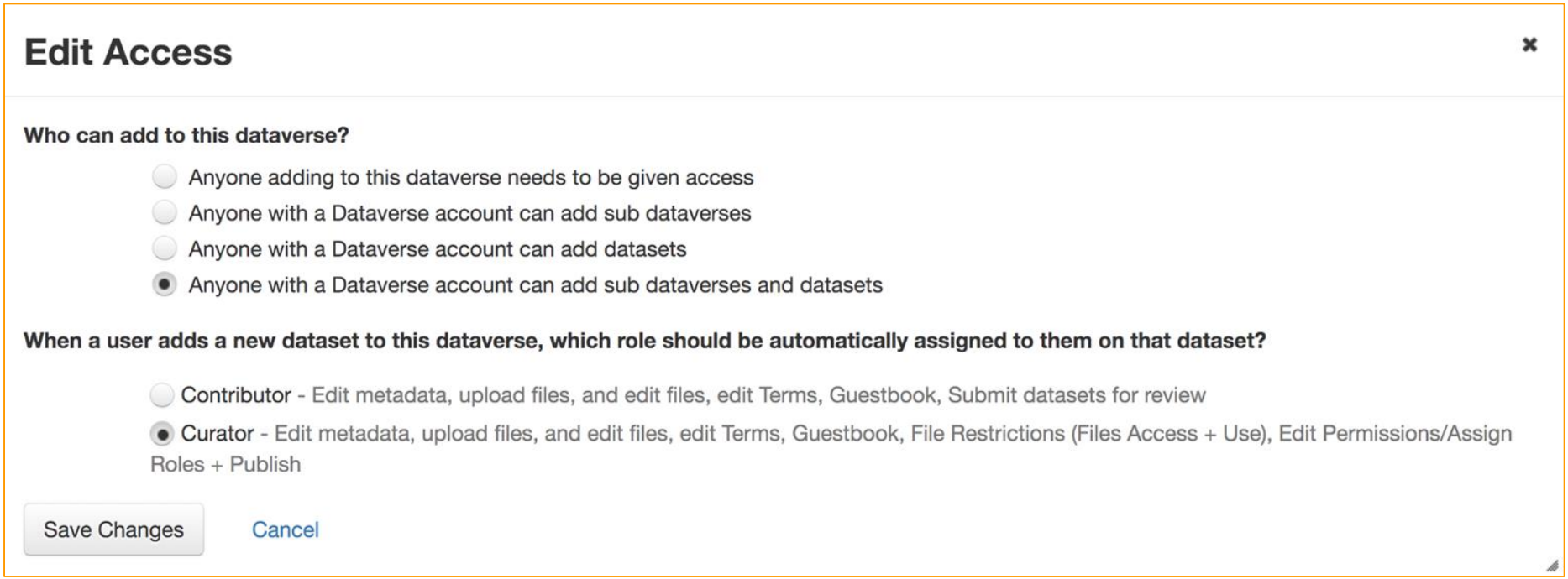
# Important Features that support data sharing

- Support for FAIR Data Principles
- Data citation for datasets and files
- OAI-PMH (Harvesting)
- APIs for interoperability and custom integrations
- Login via Shibboleth
- Login via ORCID, Google, or GitHub
- DataCite integration
- Usage statistics and metrics
- Schema.org JSON-LD
- Preview and analysis of tabular files
- External Tools
- Fixity checks for files
- Publishing workflow support
- File download in R and TSV format
- Versioning
- Custom Terms of Use
- Guestbook
- Faceted search
- File hierarchy
- Restricted files
- Customization of Dataverse Collections
- Dropbox integration
- Notifications
- Widgets
- User management
- Backend storage on S3 or Swift
- Mapping of geospatial files
- Handling large data
- Pull header metadata from Astronomy (FITS) files

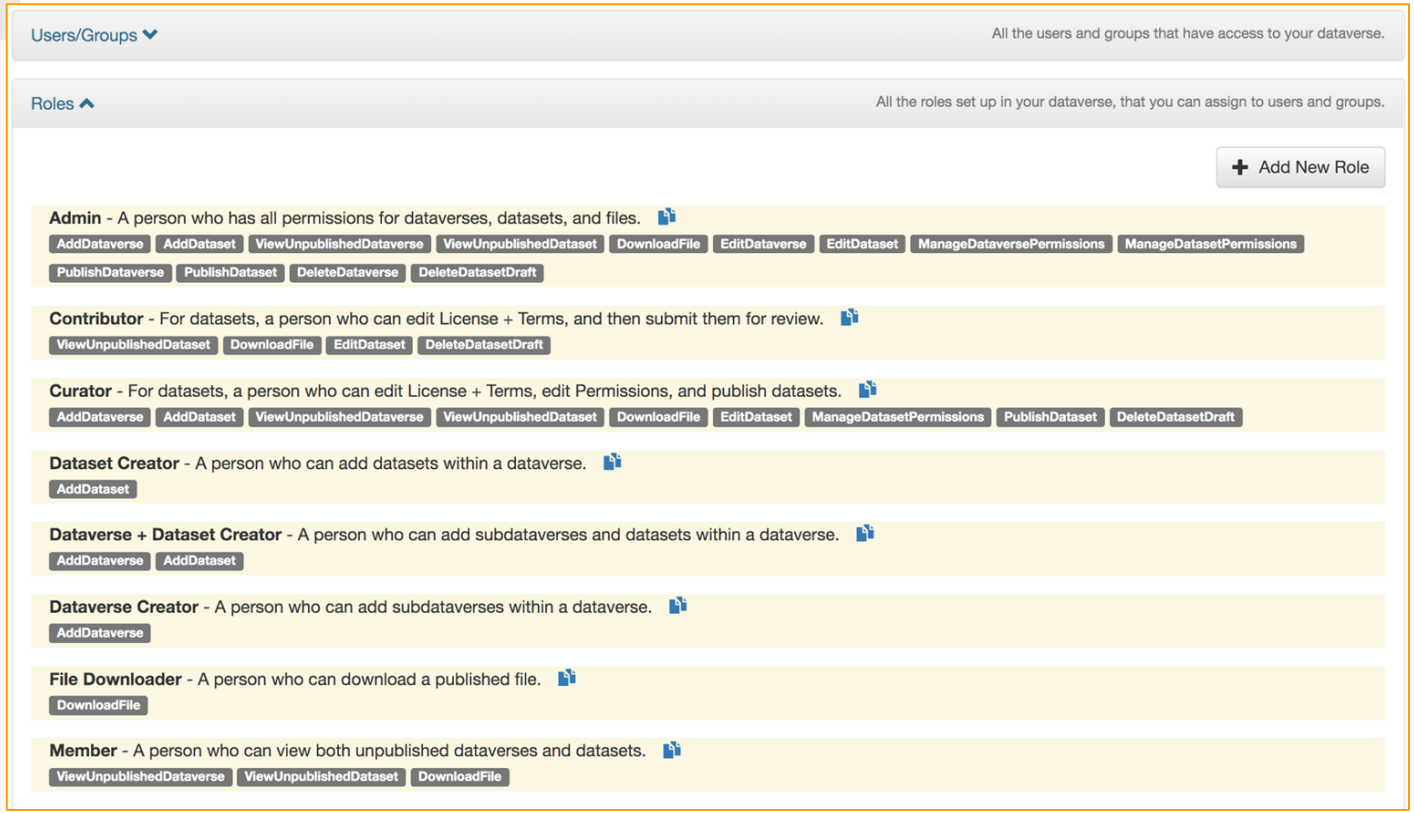
https://dataverse.org/software-features


A Dataverse Installation
Dataverse Collections
Datasets
Collection
Collection
Collection
Files
Meta data
Terms
versions

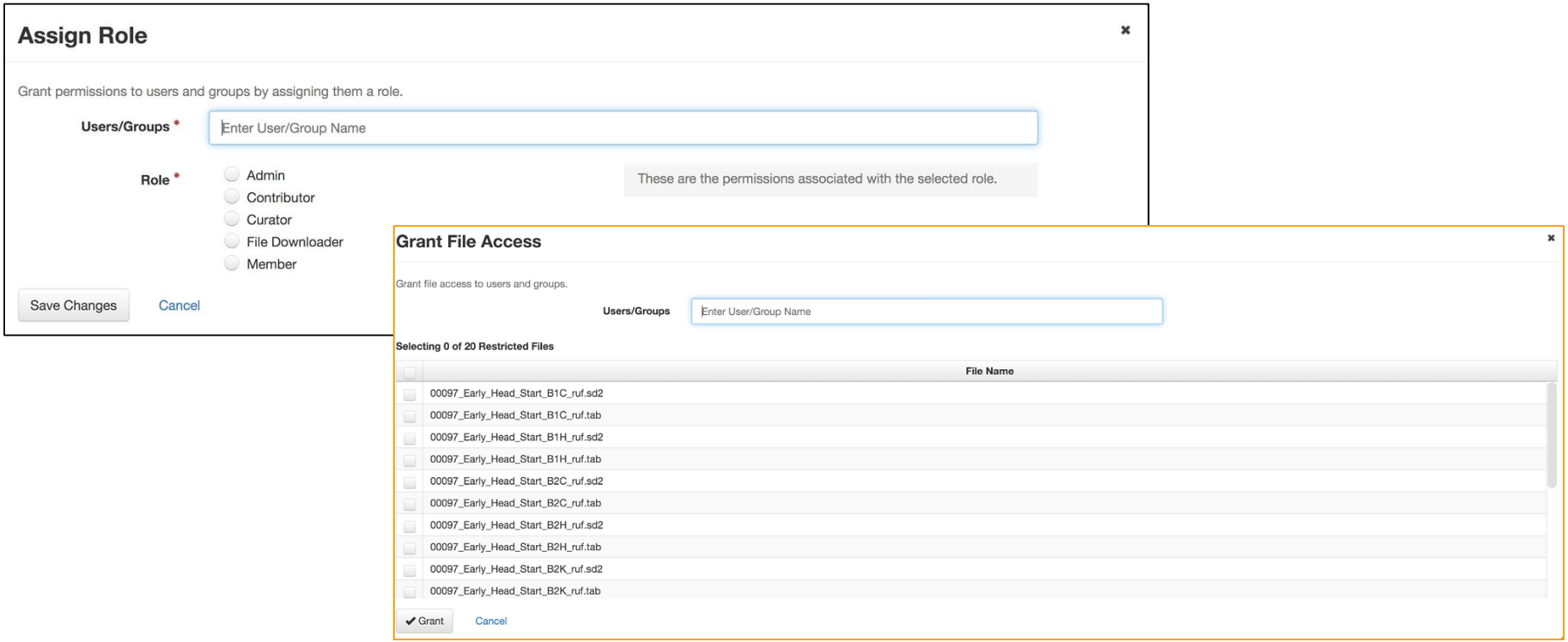
# Permissions, workflows and roles



# Permissions, workflows and roles



# Permissions, workflows and roles



# Today’s Agenda

**What is The Dataverse Project?**

**The Harvard Dataverse Repository**

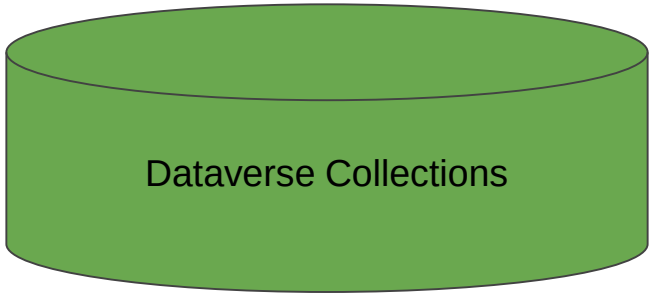
**How the Dataverse software supports research data sharing**

## Dataverse collections, datasets, and files

Dataverse and FAIR data principles

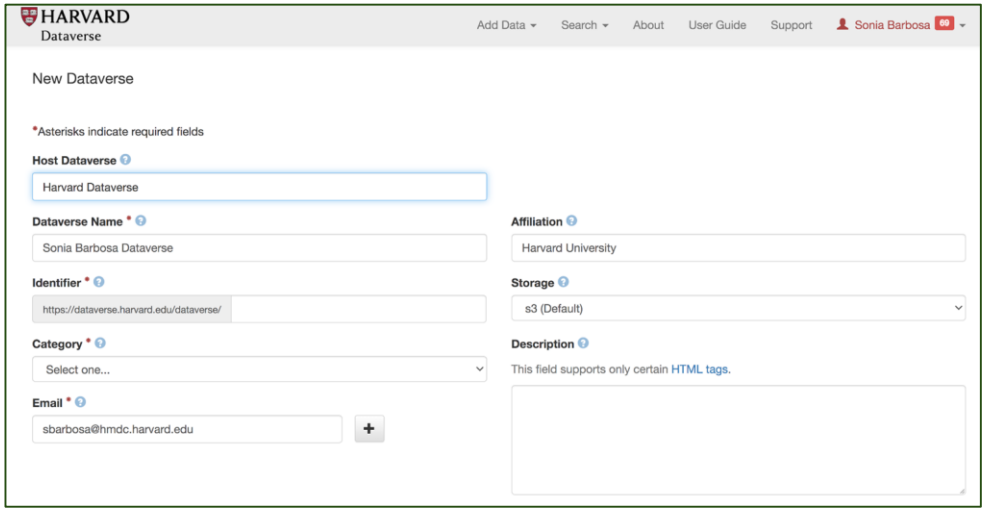
A Dataverse Installation

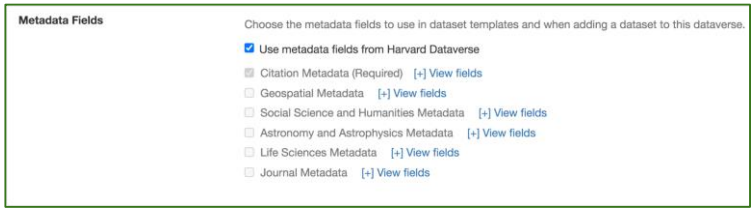
- Development Team
- Curation Team
- User support
- Institutional/Organization data repository
- Governance, preservation, stewardship, funding
- What data will you support?
- How much data can you support?
- Institutional collaborations (Library, IT, Sponsored research, etc...)
- Who will curate your content? (self-curated (Harvard Dataverse), staff curated, mixed method? etc...)
- Will you support integrated tools? (Data explorer (visualization, analysis), DataTags (sensitive data), etc...)
- Data management policies, best practices
- **GDCC member**

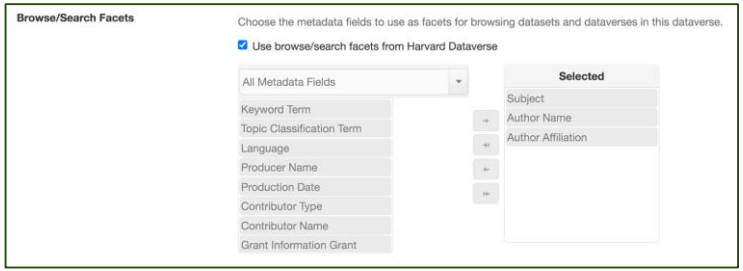


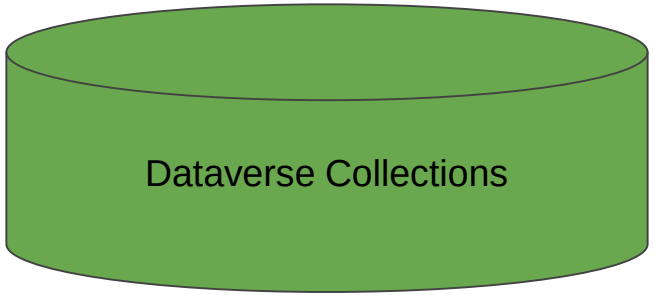

- Metadata
- Searchable facets
- Holds datasets

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

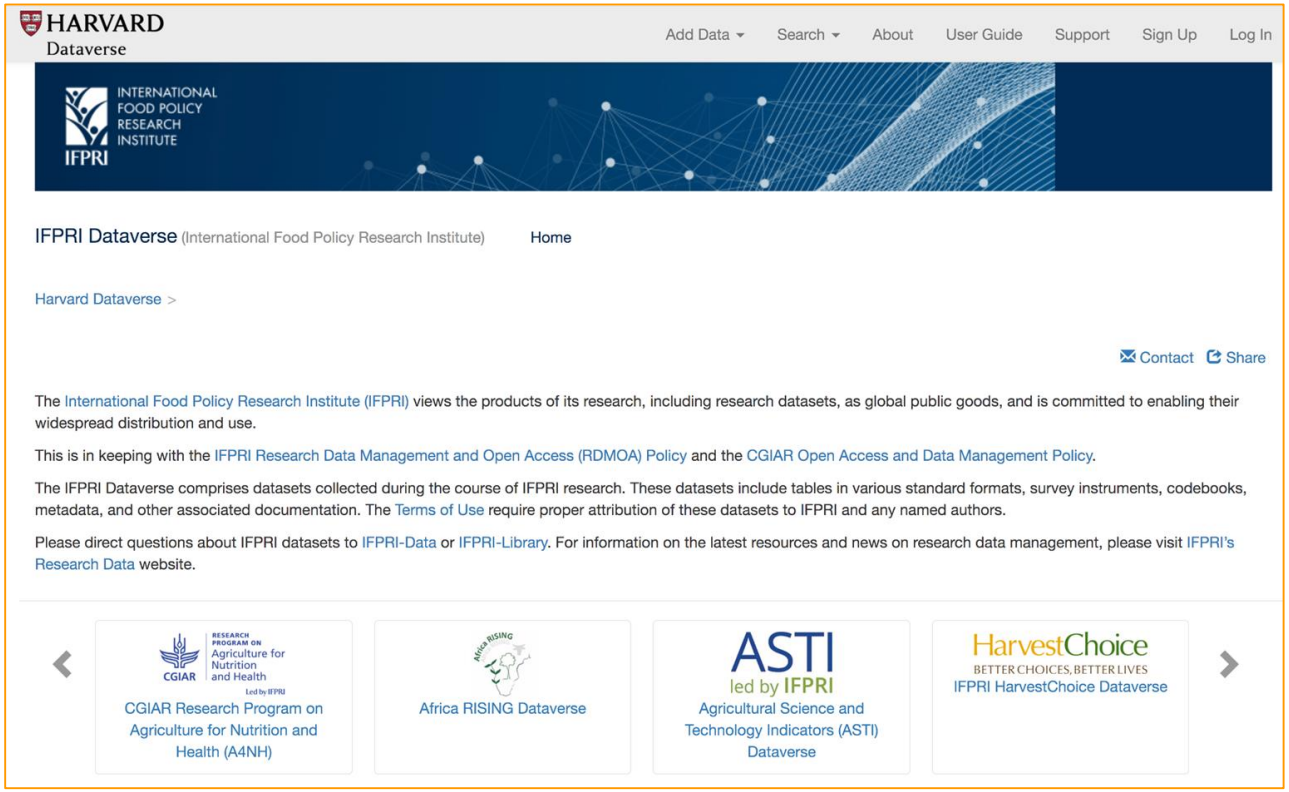





- General Information - edit your metadata, facets, description
- Theme + Widgets - customize, add widget to your website
- Permissions
- Groups
- Dataset Templates - customize consistent data deposits, terms, etc...
- Dataset Guestbooks - who is using your data and why?
- Featured Dataverses - highlight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




Dataverses (9)
Datasets (602)
Files (11,208)

**Dataverse Category**
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...

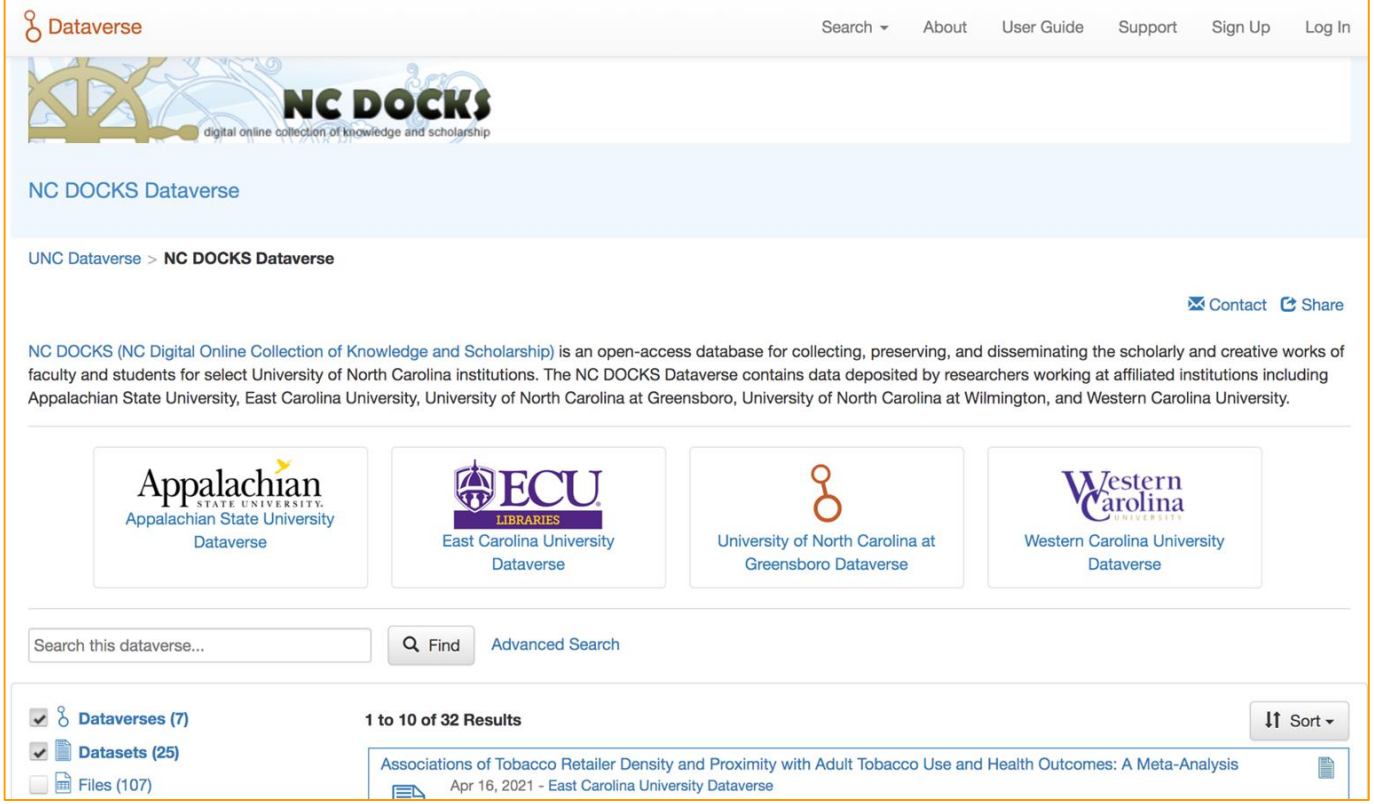
**Keyword Term**
AFRICA (215)
AFRICA SOUTH OF SAHARA (203)
EAST AFRICA (114)
nutrition (98)
Africa (96)
More...

**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**

https://dataverse.harvard.edu/dataverse/IFPRI/




Dataverses (7)
Datasets (25)
Files (107)

**Dataverse Category**
Organization or Institution (2)
Researcher (2)

**Publication Year**
2013 (5)
2016 (5)
2019 (5)
2020 (5)
2012 (3)
More...

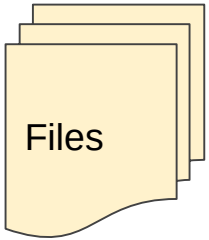
**Author Name**
Lee, Joseph (9)
Averett, Paige (3)
Cohen, Dale (2)
Henriksen, Lisa (2)
Li, Zhijin (2)
More...

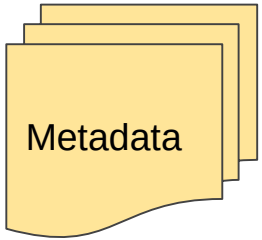
**Subject**
Medicine, Health and Life Sciences (10)
Social Sciences (9)
Earth and Environmental Sciences (6)

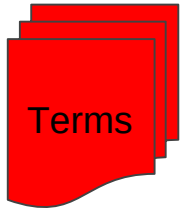
**Production Date**
2012 (4)
2013 (2)
2014 (2)
2016 (2)
1988 (1)
More...

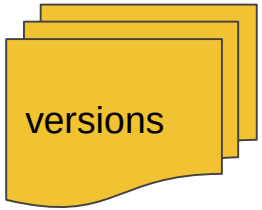
https://dataverse.unc.edu/dataverse/ncdocks

Datasets

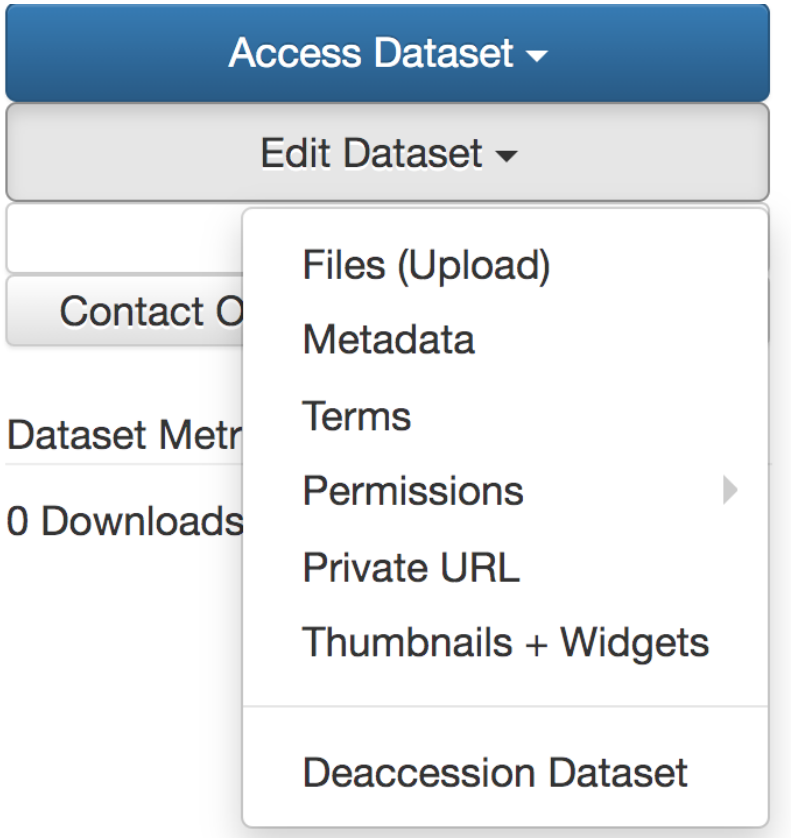

Files


Metadata


Terms


versions

Datasets


Access Dataset
Edit Dataset
Files (Upload)
Metadata
Terms
Permissions
Private URL
Thumbnails + Widgets
Deaccession Dataset
Contact O
Dataset Metr
0 Downloads

# Datasets

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

Jun 8, 2021

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

# Datasets - citations

# Datasets

**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

Read full Description [+]

**Subject** Agricultural Sciences

**Keyword** land degradation, sustainable agriculture, soil erosion, TANZANIA, EAST AFRICA, AFRICA SOUTH OF SAHARA, AFRICA

**Notes** If you are interested in accessing restricted datasets, kindly fill out this data use agreement form.

Files | Metadata | Terms | Versions

Search this dataset...

Filter by

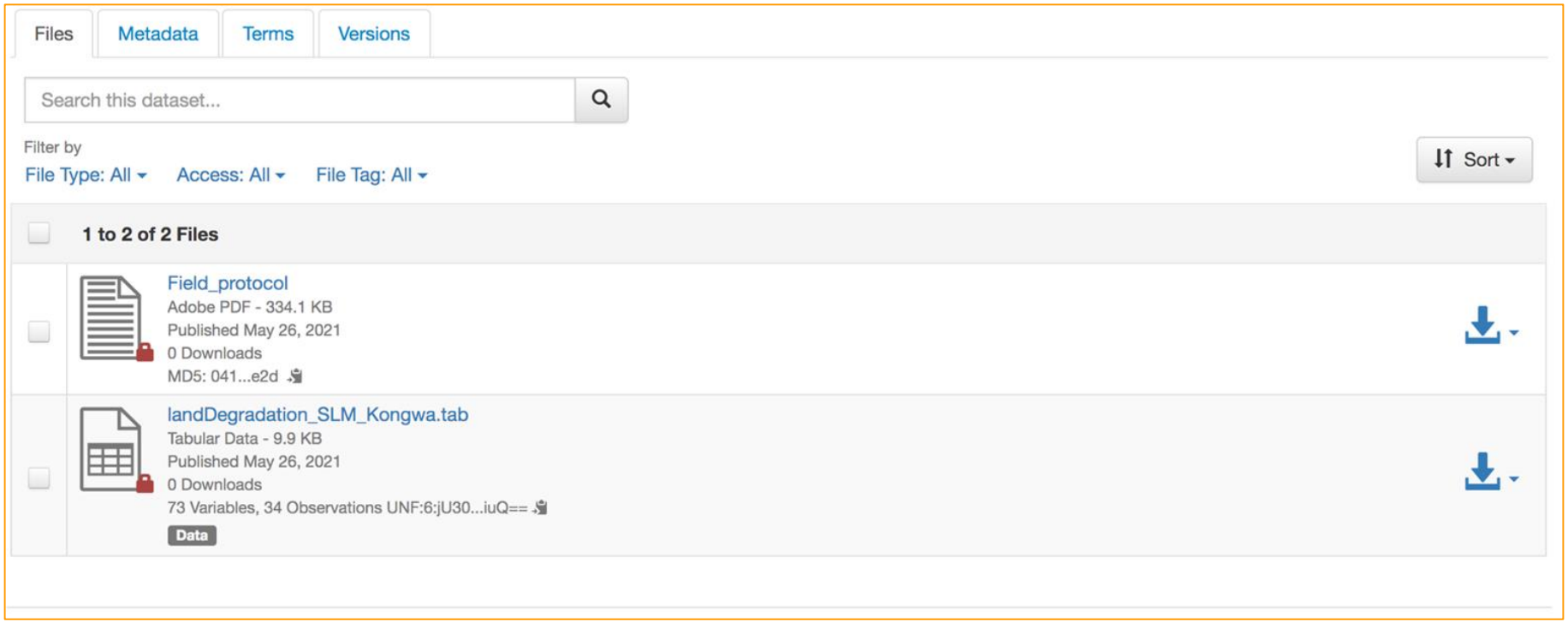
File Type: All | Access: All | File Tag: All

Sort

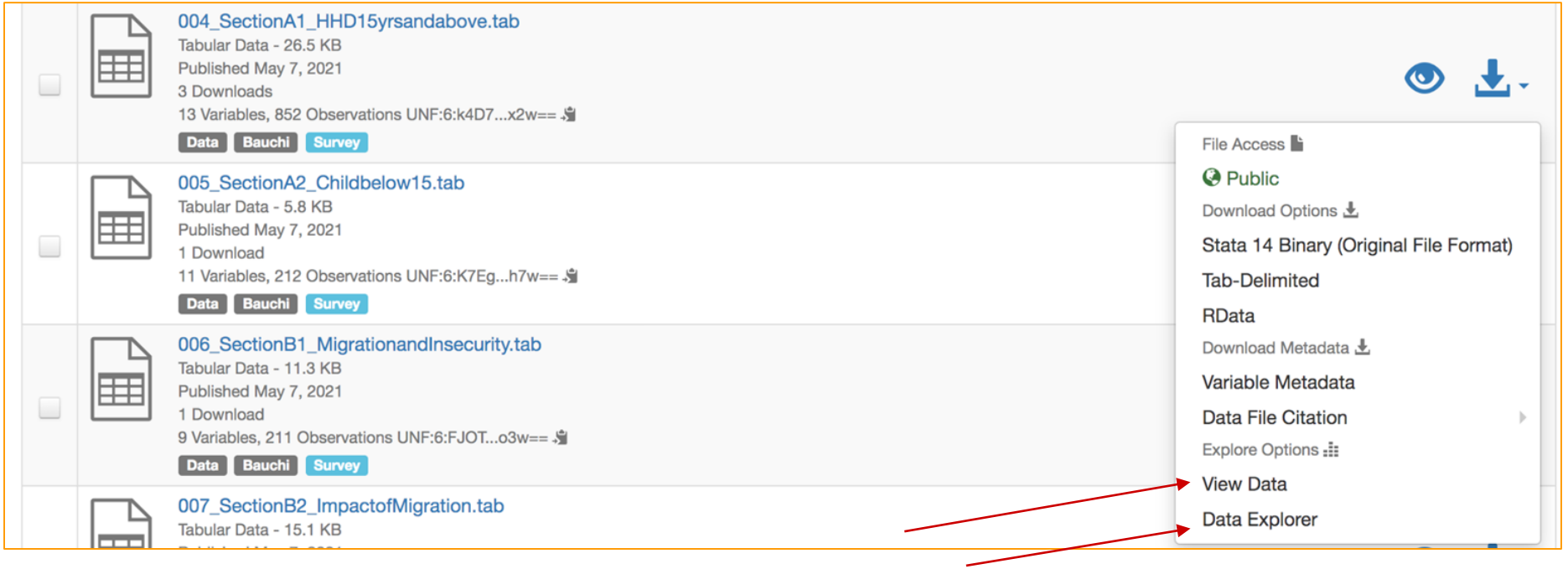
**1 to 2 of 2 Files**

Field_protocol
Adobe PDF - 334.1 KB
Published May 26, 2021
0 Downloads
MD5: 041...e2d

# Datasets - files



# Datasets - tabular files




https://guides.dataverse.org/en/latest/user/tabulardataingest/index.html

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.

# Datasets - tabular files

**Third National Fadama Development Financing II Impact Study Household Survey in Bauchi**

**004_SectionA1_HHD15yrsandabove.tab**

International Food Policy Research Institute (IFPRI), 2021, "Third National Fadama Development Financing II Impact Study Household Survey in Bauchi", https://doi.org/10.7910/DVN/AEROHZ, Harvard Dataverse, V1, UNF:6:fU65LD7Bc80cPPs7DKCBHQ== [fileUNF]

13 Results | Download

| ID | Name | Label | C |
|---|---|---|---|
| 22529304 | householdid | household ID | |
| 22529305 | memname | Name of HH member who is above 15 years old | |
| 22529302 | age | Age of member | |
| 22529309 | reltohhh | Relationship to household head | 7 |
| 22529307 | otherreltohh | Other relationship to household head (specify) | |
| 22529311 | gender | Gender of member | 2 |
| 22529301 | educlvl | Highest Level of education | 9 |
| 22529299 | othereduclvl | Other level of education(specify) | |
| 22529303 | priactvty16 | Primary activity in 2016 | 2 |
| 22529308 | othpriactvty16 | Other primary activity(specify)in 2016 | |

Chart View | Table View

**Variable educlvl: Highest Level of education**

| Values | Categories |
|---|---|
| 8 | Koranic education |
| 5 | Some secondary education (Incl. Junior secondary school) |
| 2 | Adult literacy training |
| 1 | No formal education |
| 99 | Others |
| 7 | Post-secondary education |
| 6 | Completed secondary education |
| 3 | Some primary education |
| 4 | Completed primary education |

Summary Statistics

**Variable othpriactvty16: Other primary activity(specify)in 2016**

| Values | Categories |
|---|---|

Summary Statistics

**Variable priactvty16: Primary activity in 2016**

| Values | Categories |
|---|---|
| 14 | Student in school (any type) |
| 2 | Livestock production |
| 7 | Transportation business |
| 10 | Construction |
| 18 | Artisians(Incl. Mechanics) |
| 3 | Fisheries |
| 12 | Public sector employment |
| 13 | Domestic duties |
| 4 | Forest production and/or harvesting |
| 15 | Retired |

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 | othereduclvl | 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 |

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

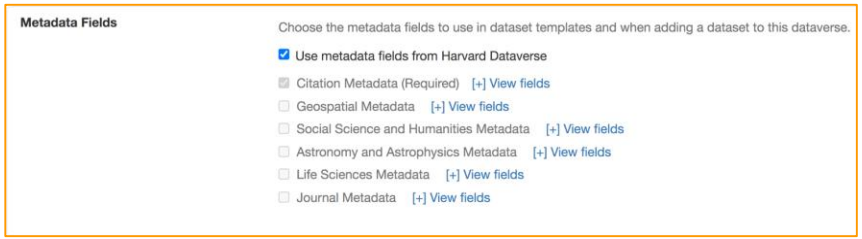
# Datasets - metadata

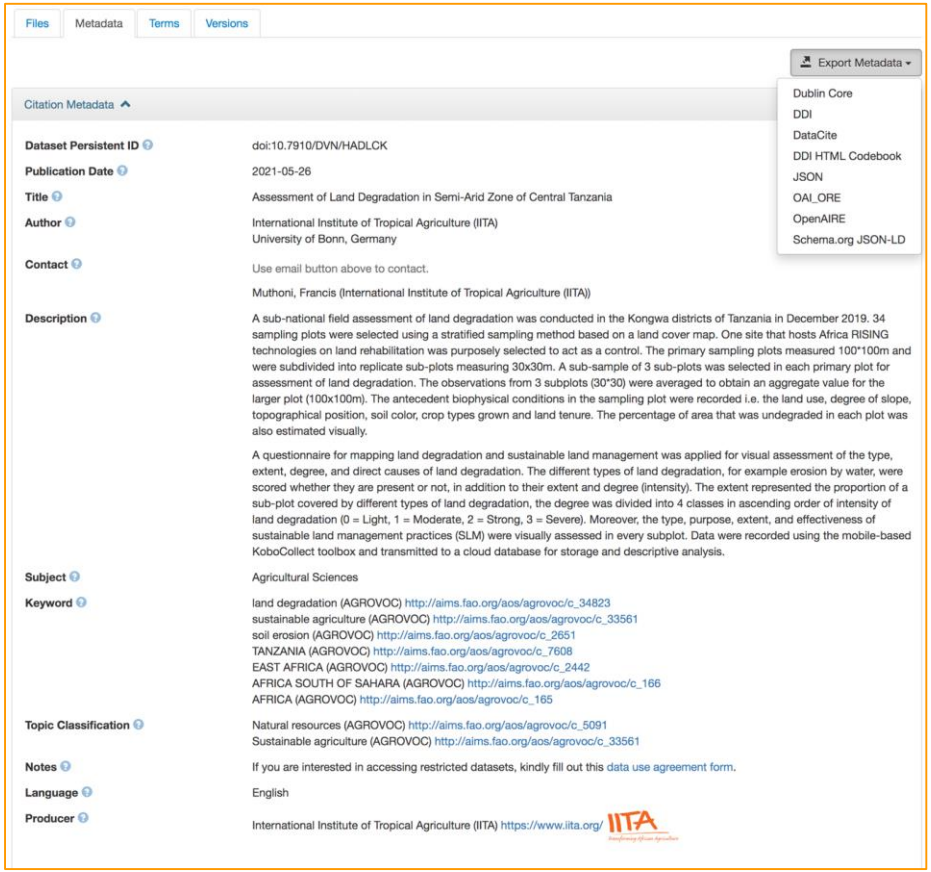
Detailed below are what metadata schemas we support for Citation and Domain Specific Metadata in the Dataverse Project:

- 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).

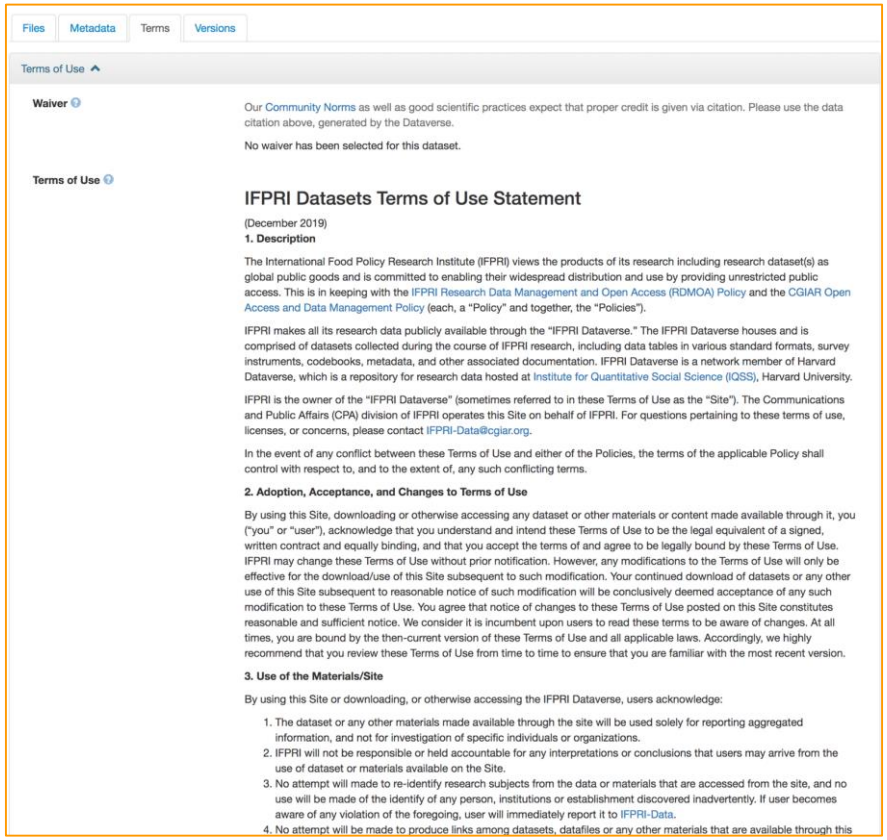
See also the Dataverse Software 4.0 Metadata Crosswalk: DDI, DataCite, DC, DCTerms, VO, ISA-Tab document and the Metadata Customization section of the Admin Guide.

# Datasets - metadata



# Datasets - terms



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.

No waiver has been selected for this dataset.

**Terms of Use**

## IFPRI Datasets Terms of Use Statement

(December 2019)

**1. Description**

The International Food Policy Research Institute (IFPRI) views the products of its research including research dataset(s) as global public goods and is committed to enabling their widespread distribution and use by providing unrestricted public access. This is in keeping with the IFPRI Research Data Management and Open Access (RDMOA) Policy and the CGIAR Open Access and Data Management Policy (each, a "Policy" and together, the "Policies").

IFPRI makes all its research data publicly available through the "IFPRI Dataverse." The IFPRI Dataverse houses and is comprised of datasets collected during the course of IFPRI research, including data tables in various standard formats, survey instruments, codebooks, metadata, and other associated documentation. IFPRI Dataverse is a network member of Harvard Dataverse, which is a repository for research data hosted at Institute for Quantitative Social Science (IQSS), Harvard University.

IFPRI is the owner of the "IFPRI Dataverse" (sometimes referred to in these Terms of Use as the "Site"). The Communications and Public Affairs (CPA) division of IFPRI operates this Site on behalf of IFPRI. For questions pertaining to these terms of use, licenses, or concerns, please contact IFPRI-Data@cgiar.org.

In the event of any conflict between these Terms of Use and either of the Policies, the terms of the applicable Policy shall control with respect to, and to the extent of, any such conflicting terms.

**2. Adoption, Acceptance, and Changes to Terms of Use**

By using this Site, downloading or otherwise accessing any dataset or other materials or content made available through it, you ("you" or "user"), acknowledge that you understand and intend these Terms of Use to be the legal equivalent of a signed, written contract and equally binding, and that you accept the terms of and agree to be legally bound by these Terms of Use. IFPRI may change these Terms of Use without prior notification. However, any modifications to the Terms of Use will only be effective for the download/use of this Site subsequent to such modification. Your continued download of datasets or any other use of this Site subsequent to reasonable notice of such modification will be conclusively deemed acceptance of any such modification to these Terms of Use. You agree that notice of changes to these Terms of Use posted on this Site constitutes reasonable and sufficient notice. We consider it is incumbent upon users to read these terms to be aware of changes. At all times, you are bound by the then-current version of these Terms of Use and all applicable laws. Accordingly, we highly recommend that you review these Terms of Use from time to time to ensure that you are familiar with the most recent version.

**3. Use of the Materials/Site**

By using this Site or downloading, or otherwise accessing the IFPRI Dataverse, users acknowledge:

1. The dataset or any other materials made available through the site will be used solely for reporting aggregated information, and not for investigation of specific individuals or organizations.
2. IFPRI will not be responsible or held accountable for any interpretations or conclusions that users may arrive from the use of dataset or materials available on the Site.
3. No attempt will made to re-identify research subjects from the data or materials that are accessed from the site, and no use will be made of the identify of any person, institutions or establishment discovered inadvertently. If user becomes aware of any violation of the foregoing, user will immediately report it to IFPRI-Data.
4. No attempt will be made to produce links among datasets, datafiles or any other materials that are available through this

# Datasets - versions

Files | Metadata | Terms | Versions

| Dataset | Summary | Contributors | Published |
|---|---|---|---|
| 1.0 | This is the first published version. | IFPRI KM | May 26, 2021 |

https://dataverse.org/files/dataverseorg/files/metadataprovenance-mercecrosas.pdf

# Today’s Agenda

What is The Dataverse Project?

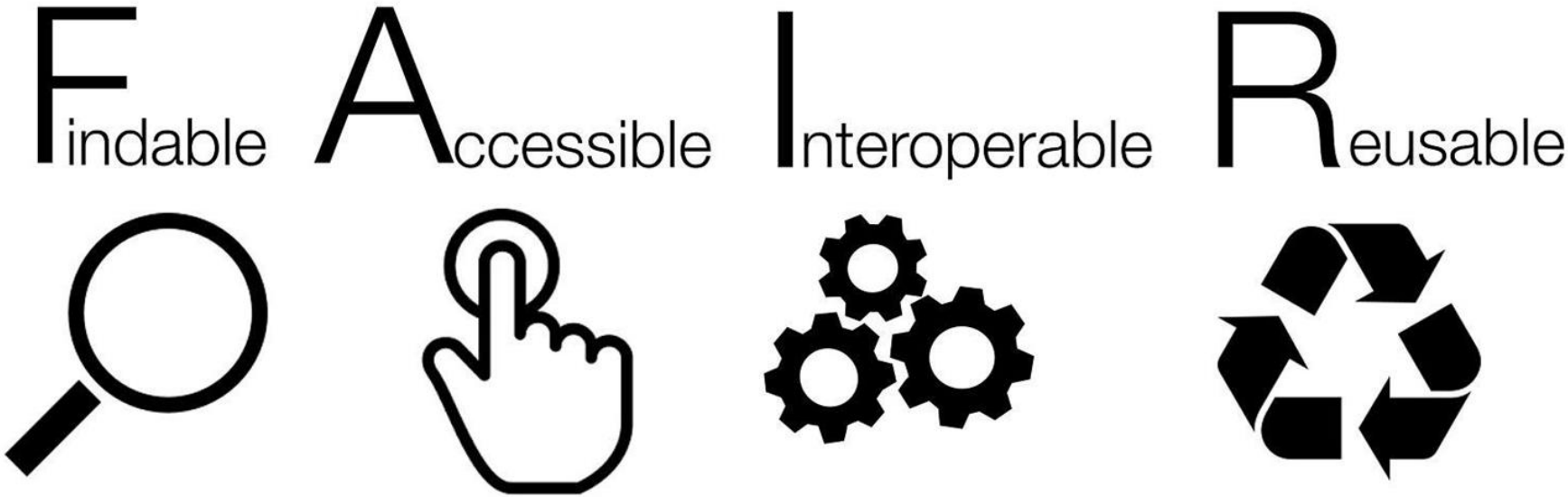
The Harvard Dataverse Repository

How the Dataverse software supports research data sharing

Dataverse collections, datasets, and files

Dataverse and FAIR data principles

# FAIR principles and Dataverse




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

# The Dataverse Project RoadMap

The Strategic Goals of the Dataverse Project are our highest-level guide.  These goals are to:

- increase adoption (users, Dataverse repositories, Dataverse collections, datasets, journals)
- develop capability to handle sensitive, large scale, and streaming data
- expand data and metadata features for existing and new disciplines
- expand archival and preservation features
- increase interoperability through implementation of standards
- increase contributions from the open-source development community
- improve UX and UI
- continue to increase the quality of the software

# THANK YOU!

support@dataverse.harvard.edu