



Dataverse Repository Training

Experience sharing

Yetimwork Habte

Organized by
IFPRI-ESSP and EPHI-NIPN

March 12, 2019

Content of the training

Will introduce you to the basics of Dataverse including:

- What is dataverse
- Preparing your submissions
- Creating account and logging in
- Accessing Dataverse pages
- How to add a dataset to your project's Dataverse page
- Details about what metadata to include
- Permissions

What is dataverse

- Dataverse is a free, credible, open access data repository system.
- It is a container for datasets and documents.
- Link here: <https://dataverse.harvard.edu/>
- Created and managed by Harvard University
- Dataverse is used by research institutions world wide
- Provides an academic data citation for all published datasets
- Improves public accessibility of datasets
- promotes collaboration and reduces the number of emails sent!
- It is also possible to privately share information with access/permission setting (setting up private url).

Step 1: Prepare Your Submission and convert it in to required format

1.1. Data cleaning

Example: data cleaning

- We have 5 survey datasets with many sections (and variables)
- Took me about 1 month/ more
- Check for incorrect data entry and missing values
- String vs numeric values (destring command or generate new numeric variable)
- Renaming and labeling (rename var, label var)
- Generating new variables (eg unique id, hh_size) (gen command)

1.2. Preparing codebook

- Is another time taking task
 - Example Preparing codebook

Variable name	variable label	data type	value	value label	Comments
hhq_id	Household ID	numeric			
zone	Zone Code	numeric	1	North Shewa	
			2	West Shewa	
			3	Sub-urban	
			4	Addis	
sectiona_1	Id code of person living in the HH	numeric			
sectiona_5	Is this person a minda	numeric	1	Yes	Minda is someone who works for the household on agriculture or livestock and lives in the household
			2	No	
sectiona_20	Do you fast	numeric	1	Yes	
			2	No	
sectiona_21	Name of fasting period observed	text			
hh_size	Household size	numeric			This is generated by counting no. of observation per HH

Step 1..... Prepare your submission

1.3. Sampling strategy (update)

1.4. Survey questionnaire

1.5. Consent form (language used)

1.6. Other important documents (eg related publications and reports)

- Once cleaned and stripped of personally identifiable information (PII), you should convert the data in to required format.
 - E.g STATA to excel (csv format/ comma delimited format)
- The codebook, sampling strategy, survey questionnaire and consent form should also be converted in to required format

Step 2: Create an Account

HARVARD
Dataverse

Harvard Dataverse > Create Account

Account Information

Username Create a valid username of 2 to 60 characters in length containing letters (a-Z), numbers (0-9), dashes (-), underscores (_), and periods (.).

Password Your password must contain:

- At least 6 characters (passwords of at least 20 characters are exempt from all other requirements)
- At least 1 character from each of the following types: letter, numeral

Retype Password

Given Name

Family Name

Email

Affiliation

Position

General Terms of Use

Harvard Dataverse General Terms of Use
Acceptance of Terms

The following terms and conditions govern all use of the Harvard Dataverse application (the Site) and the services available on or at the Site (the Site and the services taken together, the Service) by you and any third parties who use your account. Use of the Service is offered subject to acceptance of and compliance with all of the terms and conditions contained herein (the General Terms of Use) and all other applicable operating rules, policies and procedures of the Service. By clicking "I accept" and/or by your use of the Service, you agree and accept all the terms in this General Terms of Use document

☐ I have read and accept the Dataverse General Terms of Use as outlined above.

Create Account **Cancel**

1. Click here to see the sign up page

2. Fill in the required information and create your Dataverse account

Step 3: Log In

Log In

 Log in or sign up with your institutional account — [learn more](#). Leaving your institution? Please contact [Harvard Dataverse Support](#) for assistance.

Your Institution

Please select...

Continue

[Allow me to type the name of my institution](#)

Other options

Username/Email

ORCID

GitHub

Google

Step 4: Find your Institution's Dataverse Page

HARVARD
Dataverse

Metrics 3,836,677 Downloads

Contact Share

Ethiopian Public Health Institute (EPHI) Find Advanced Search + Add Data

Ethiopian Public Health Institute (EPHI)

1 to 10 of 7,594 Results

Ethiopian Public Health Institute (EPHI)

Legume Innovation Lab Dataverse (Michigan State University)
Feb 20, 2018 MSU-AFRE Food Security Group
Legume Innovation Lab Impact Assessment Project

Innovation Platforms in Dryland Systems in Central Asia.pdf
Jul 7, 2015 - Summary Reports
Adobe PDF - 845.2 KB - MD5: e3b6d3bddd163583e6806d178a77053e

Feed the Future Innovation Lab for Collaborative Research on Sustainable Intensification (SIIL) (Kansas State University)
Nov 7, 2017
The Feed the Future Innovation Lab for Collaborative Research on Sustainable Intensification (SIIL), along with the associated Geospatial and Farming Systems Research Consortium and the Appropriate Scale Mechanization Consortium, will produce measurable impacts on reducing global hunger, poverty and improving

Design and Pilot Processes for Enhancing Facilitation, Communication and Coordination of Innovation Platforms
Jan 23, 2018 - Africa RISING Dataverse
International Livestock Research Institute (ILRI), 2015, "Design and Pilot Processes for Enhancing Facilitation, Communication and Coordination of Innovation Platforms", <https://doi.org/10.7910/DVN/ZU5J0P>, Harvard Dataverse, V1
To learn more about this dataset, please visit to ILRI Datasets Portal.
Contact Affiliation: International Livestock Research Institute (ILRI)
Author Name: International Livestock Research Institute (ILRI)


Policy, Impact Assessment and Innovation Systems (AfricaRice)
Jun 20, 2017 AfricaRice Dataverse

Search Results Summary:

- Dataverses (105)**
- Datasets (6,237)**
- Files (1,252)**
- Publication Status**
 - Published (7,589)
 - Unpublished (21)
- Dataverse Category**
 - Research Project (43)
 - Research Group (18)
 - Organization or Institution (12)
 - Researcher (9)
 - Category (4)
- Data Source**
 - Harvard Dataverse (5,298)
 - Other Dataverse (2,296)
- Publication Date**
 - 2018 (3,068)
 - 2017 (1,586)
 - 2016 (832)
 - 2015 (414)
 - 2014 (313)
- Subject**
 - Social Sciences (682)
 - Agricultural Sciences (254)
 - Earth and Environmental Sciences (142)
 - Medicine, Health and Life Sciences (108)
 - Arts and Humanities (64)

Type e.g
"EPHI" in the
search bar to get
your institution's
page as the first
result, it will
show up in the
results list once
it is created.

Step 5: Find Your Project's Dataverse Page


 **HARVARD**
Dataverse


Q About User Guide Support Sign Up Log In


Harvard Dataverse > Ethiopian Public Health Institute

Contact Share

Search this dataverse... Find Advanced Search

☒  **Dataverses (4)**

☒  **Datasets (0)**

☐  **Files (0)**


Dataverse Category
Research Project (4)


Publication Date
2018 (4)


1 to 4 of 4 Results

Sort

The National Information Platform for Nutrition (NIPN)

 Empowerment of Village Women for Detection and Control of Livestock Diseases in Nepal
Jul 3, 2018

 Improving Dairy Animal Productivity and Income of Dairy Farmers through Effective Control of Mastitis Disease
Jul 3, 2018

 Feeding Support Tool Development for Enhancing Dairy Animal Productivity for Improved Livelihood of Smallholder Dairy Farmers in Nepal
Jul 3, 2018

Next, click on the name of your project

■ National Information Platform for Nutrition (NIPN)

.. [Harvard Dataverse](#) > [Ethiopian Public Health Institute](#) > [National Information Platform for Nutrition \(NIPN\)](#)

[Contact](#) [Share](#) [Edit](#) ▾

Search this dataverse...

 Find

[Advanced Search](#)

☒  **Dataverses (0)**

☒  **Datasets (0)**

☐  **Files (0)**

This dataverse currently has no dataverses, datasets, or files. You can add to it by using the Add Data button

+ Add Data ▾

New Dataverse

New Dataset

Click on the “add data” button and select “new dataset”

Step 6: Fill in the Metadata (provide information about dataset(s))

Add the title of the dataset
E.g. Data for child nutrition survey

National Information Platform for Nutrition (NIPN)

Harvard Dataverse > Ethiopian Public Health Institute > National Information Platform for Nutrition (NIPN)

Host Dataverse National Information Platform for Nutrition (NIPN)

Citation Metadata

Title *

Enter title...

Add "Replication Data for" to Title

Author *

Name *

FamilyName, GivenName or Organization

Affiliation

Identifier Scheme

Select...

Identifier

Contact *

Name

FamilyName, GivenName or Organization

Affiliation

E-mail

Description *

This field supports only certain HTML tags.

Text

Date

YYYY-MM-DD

Fill in the information of the person responsible for the dataset and the submitter (if a different person)

Add a detailed description of the dataset answering who, what, when, where, why, and how?

Add the date when the dataset was **first created**

Relevant Options: medicine health & life sciences; agricultural sciences; or other

Keywords are used by others to search for your datasets in Dataverse. Choose strategically to make the data easy to search for

Related: include the title and the link of the work produced using this data, if any.

The image shows a screenshot of the Dataverse dataset creation form. Several fields are circled in red and labeled with arrows from external text boxes:

- Subject**: A dropdown menu with a red circle around it and an arrow pointing to the 'Relevant Options' text box.
- Keyword**: A text input field with a red circle around it and an arrow pointing to the 'Keywords are used...' text box.
- Related Publication**: A text input field with a red circle around it and an arrow pointing to the 'Related:' text box.
- Notes**: A large text area with a red circle around it and an arrow pointing to the 'Notes Guidelines:' text box.
- Plus button**: A small square button with a plus sign (+) inside a red circle, with an arrow pointing to the 'Keyword Guidelines:' text box.

The form itself includes the following sections:

- Subject**: A list of checkboxes for various fields: Computer and Information Science, Engineering, Business and Management, Astronomy and Astrophysics, Medicine, Health and Life Sciences, and Earth and Environmental Sciences.
- Term**: A text input field.
- Vocabulary**: A text input field.
- Vocabulary URL**: A text input field with a placeholder 'Enter full URL, starting with http://'.
- Citation**: A large text area.
- ID Type**: A dropdown menu with 'Select...' as the current selection.
- ID Number**: A text input field.
- URL**: A text input field with a placeholder 'Enter full URL, starting with http://'.
- Depositor**: A text input field.
- Deposit Date**: A text input field with a placeholder 'YYYY-MM-DD'.

Keyword Guidelines:
Click on the plus (+) button to add more Keywords or “Terms”

Notes Guidelines:

In the “Notes” section, include answers to the following:

- 1.If this data relate to living, human subjects or not
- 2.If informed consent was obtained or not
3. summary of the data analysis done using this data (a paragraph is enough)

Step 7: Upload your Dataset

Files

File upload limit is 2,684,354,560 bytes per file. For more information about supported file formats, please refer to the [User Guide](#).

+ Select Files to Add

Drag and drop files here.

Upload from Dropbox

Files can also be uploaded directly from Dropbox.

Files you upload will appear here.

Metadata Tip: After adding the dataset, click the Edit Dataset button to add more metadata.

Save Dataset Cancel

File Upload Guidelines:

1. Datasets
 2. Codebooks
 3. Survey protocols and instruments
 4. Methodologies for data analysis
 5. Any annotations describing data quality issues in the data
 6. Consent form language
- Datasets submission format: excel(.csv), stata, spss, etc

Reminder: Be sure to save your dataset!

Tip: Metadata can be edited in Dataverse after a dataset is submitted

M&E Dataverse **Unpublished**

Harvard Dataverse > Ethiopian Public Health Institute > National Information Platform for Nutrition (NIPN)

Metrics 0 Downloads

ME internal survey 2 **Draft** **Unpublished**

Harris-Coble, Lacey, 2018, "ME internal survey 2", <https://doi.org/10.7910/DVN/RDMPWW>, Harvard Dataverse, DRAFT VERSION

Description test

Subject Other

Files Metadata Terms Versions

0 Files

There are no files in this dataset.

Upload Files

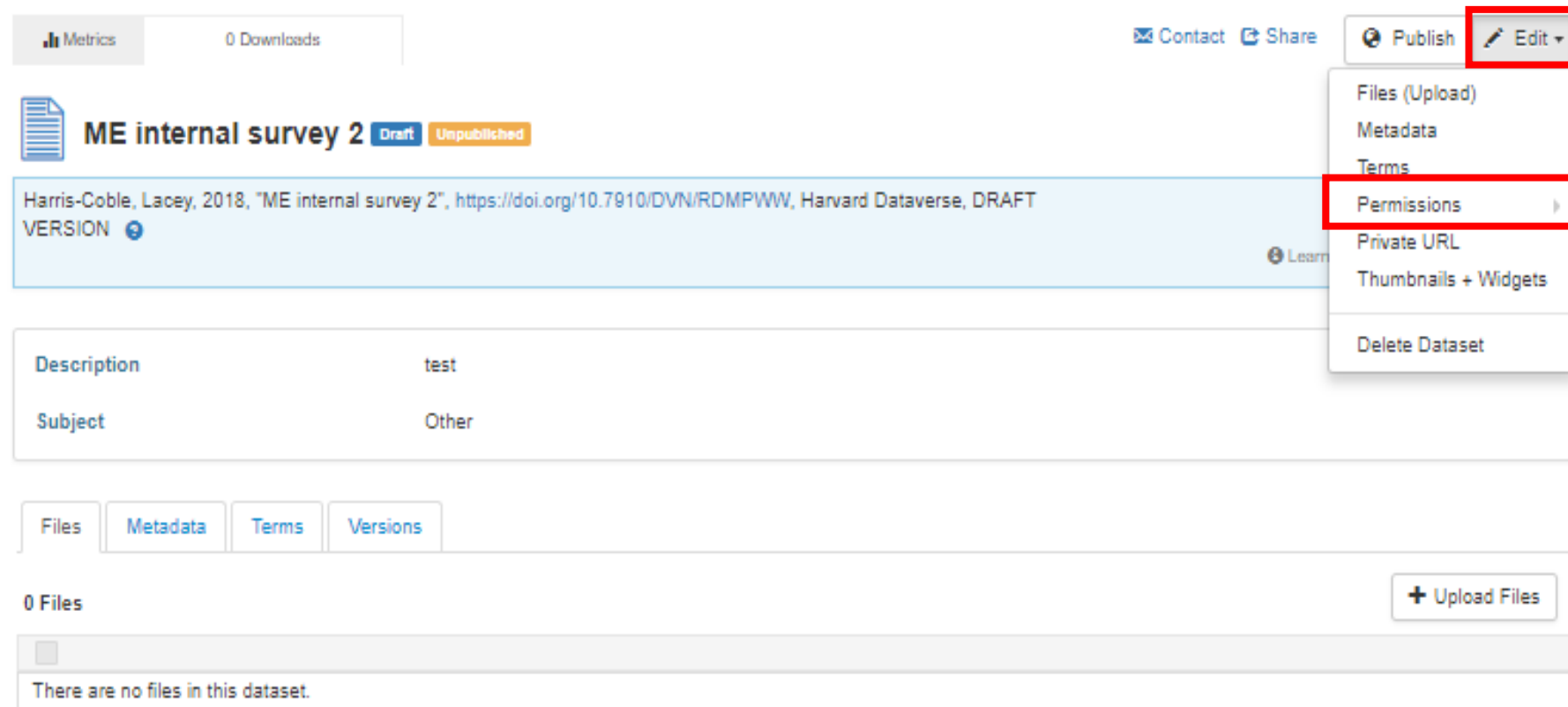
When you are ready to make the dataset published click the publish button

To add files, edit the metadata or delete the dataset, click edit and then use the drop down menu

Step 8: Permissions

M&E Dataverse **Unpublished**

Harvard Dataverse > Ethiopian Public Health Institute > National Information Platform for Nutrition (NIPN) M&E Dataverse



Metrics 0 Downloads

Contact Share Publish **Edit**

ME internal survey 2 **Draft** **Unpublished**

Harris-Coble, Lacey, 2018, "ME internal survey 2", <https://doi.org/10.7910/DVN/RDMPWW>, Harvard Dataverse, DRAFT VERSION

Description test

Subject Other

Files Metadata Terms Versions

0 Files [+ Upload Files](#)

There are no files in this dataset.

Click Edit and then Permissions to change permission settings for your project Dataverse page

M&E Dataverse Unpublished

Harvard Dataverse > The Feed the Future Innovation Lab for Livestock Systems > Livestock Lab Management Entity Dataverse > M&E Dataverse > Permissions

Permissions ⌵ ⓘ Current access configuration to your dataverse.

Select if all users or only certain users are able to add to this dataverse, by clicking the Edit Access button.

⚙️ Edit Access

National Information Platform for Nutrition (NIPN)

Harvard Dataverse > Ethiopian Public Health Institute > National Information Platform for Nutrition (NIPN) > Dataset permissions

Users/Groups ⌵ ⓘ All the users and groups that have access to your dataverse.

1 User/Group

User/Group Name (Affiliation) ⌵	ID ⌵	Role ⌵	Action
Lacey Harris-Coble	@livestock-lab	Admin	✕ Remove Assigned Role

👤 Assign Roles to Users/Groups

Roles ⌵ ⓘ All the roles set up in your dataverse, that you can assign to users and groups.

- Admin** - A person who has all permissions for dataverses, datasets, and files.
AddDataverse AddDataset ViewUnpublishedDataverse ViewUnpublishedDataset DownloadFile EditDataverse EditDataset ManageDataversePermissions ManageDatasetPermissions PublishDataverse PublishDataset DeleteDataverse DeleteDatasetDraft
- Contributor** - For datasets, a person who can edit License + Terms, and then submit them for review.
ViewUnpublishedDataset DownloadFile EditDataset DeleteDatasetDraft
- Curator** - For datasets, a person who can edit License + Terms, edit Permissions, and publish datasets.
AddDataverse AddDataset ViewUnpublishedDataverse ViewUnpublishedDataset DownloadFile EditDataset ManageDatasetPermissions PublishDataset DeleteDataset
- 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 file.
DownloadFile
- Member** - A person who can view both unpublished dataverses and datasets.
ViewUnpublishedDataverse ViewUnpublishedDataset DownloadFile

To give members of your research team the ability to add and edit datasets assign them the role of admin by clicking "Assign Roles to Users/Groups"

Assign Role

Grant permissions to users and groups by assigning them a role.

Users/Groups

Role ☐ Admin
☐ Contributor
☐ Curator
☐ Dataset Creator
☐ Dataverse + Dataset Creator
☐ Dataverse Creator
☐ File Downloader
☐ Member

These are the permissions associated with the selected user/group

Save Changes Cancel

Search for the researchers username and assign a role (keep in mind the search is case sensitive)

Admin - A person who has all permissions for dataverses, datasets, and files.

Contributor - For datasets, a person who can edit License + Terms, and then submit them for review. Can not publish dataset.

Curator - For datasets, a person who can edit License + Terms, edit Permissions, and publish datasets.

File Downloader - A person who can download a file (including unpublished files)

Member - A person who can view both unpublished dataverses and datasets, and download files..

Questions or Comments?

Thank you!