Dataverse: evoluzione, prospettive e sfide di un ecosistema aperto per l'Open Science

Susanna Mornati – 4Science

GenOA Week, 19 novembre 2025

4SCIENCE
Share your knowledge

The Dataverse Project

Cos'è Dataverse

Piattaforma open source per la gestione, pubblicazione e conservazione dei dati della ricerca in ottemperanza ai principi F.A.I.R. (findable, accessible, interoperable, reusable).

Sviluppata dall'Institute for Quantitative Social Science di Harvard, e mantenuta attraverso una community internazionale.

E' adottata a livello mondiale da università, enti di ricerca, istituzioni pubbliche e private.

Già diverse installazioni in Italia: Statale di Milano, Genova, Venezia, Bocconi, IIT

Filosofia e valori

• OPEN SOURCE

• PRINCIPI FAIR

• TRASPARENZA E COLLABORAZIONE

• COMUNITÀ SCIENTIFICA AL CENTRO

A chi è rivolto

Open source research data repository software

Researchers

Enjoy full control over your data. Receive *web visibility, academic credit,* and *increased citation counts.* A personal Dataverse collection is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. Want to set up your personal Dataverse collection?

Journals

Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data.* Participate in the open data movement by using a Dataverse collection as part of your journal data policy or list of repository recommendations. Want to find out more about journal Dataverse collections?

Institutions

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. Participate in the drive to set norms for sharing, preserving, citing, exploring, and analyzing research data. Want to install a Dataverse repository?

Developers

Participate in a vibrant and growing community that is helping to drive the norms for sharing, preserving, citing, exploring, and analyzing research data. 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. Want to contribute?

Le risorse

The Dataverse Project

Community

- Community Meetings
- Community Calls
- Global Dataverse Community Consortium

Thematic working groups

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

https://www.gdcc.io/working-groups.html

La diffusione

Supporto ai dati FAIR

How Dataverse Supports FAIR

- FINDABLE: Persistent Identifiers (DOIs, Handles) for datasets (and files), and enabling metadata indexing by search engines
- ACCESSIBLE: Open access to (meta)data and ensures data can be downloaded in machine-readable formats
- INTEROPERABLE: Provides standardised metadata schemas and enabling the integration of of data with research tools and platforms
- REUSABLE: Licenses that clearly state how data can be used and ensuring that data are well documented and preserved for long term use

Gestire e condividere

- Facilità di deposito
- Formato di citazione
- Controllo degli accessi
- Standard di metadati
- Politiche di condivisione
- Funzioni di discovery

Supported metadata schemas (compliant with DDI Lite, DDI 2.5 Codebook, DataCite 3.1, Dublin Core's DCMI Metadata Terms)

- Citation Metadata. Language field uses ISO 639-1.
- Geospatial Metadata. Country/Nation field uses ISO 3166-1.
- Social Science & Humanities Metadata.
- Astronomy and Astrophysics Metadata, can be mapped to IVOA VOResource Schema format.
- Life Sciences Metadata: based on ISA-Tab Specification, OBI Ontology, NCBI Taxonomy for Organisms.
- Journal Metadata: based on the Journal Archiving and Interchange Tag Set, version 1.2.
- CodeMeta Software Metadata, version 2.0.
- Computational Workflow Metadata, adapted from Bioschemas Computational Workflow Profile, v. 1.0 and Codemeta.

Funzionalità

Dataverse Repository Features

The Dataverse® Project

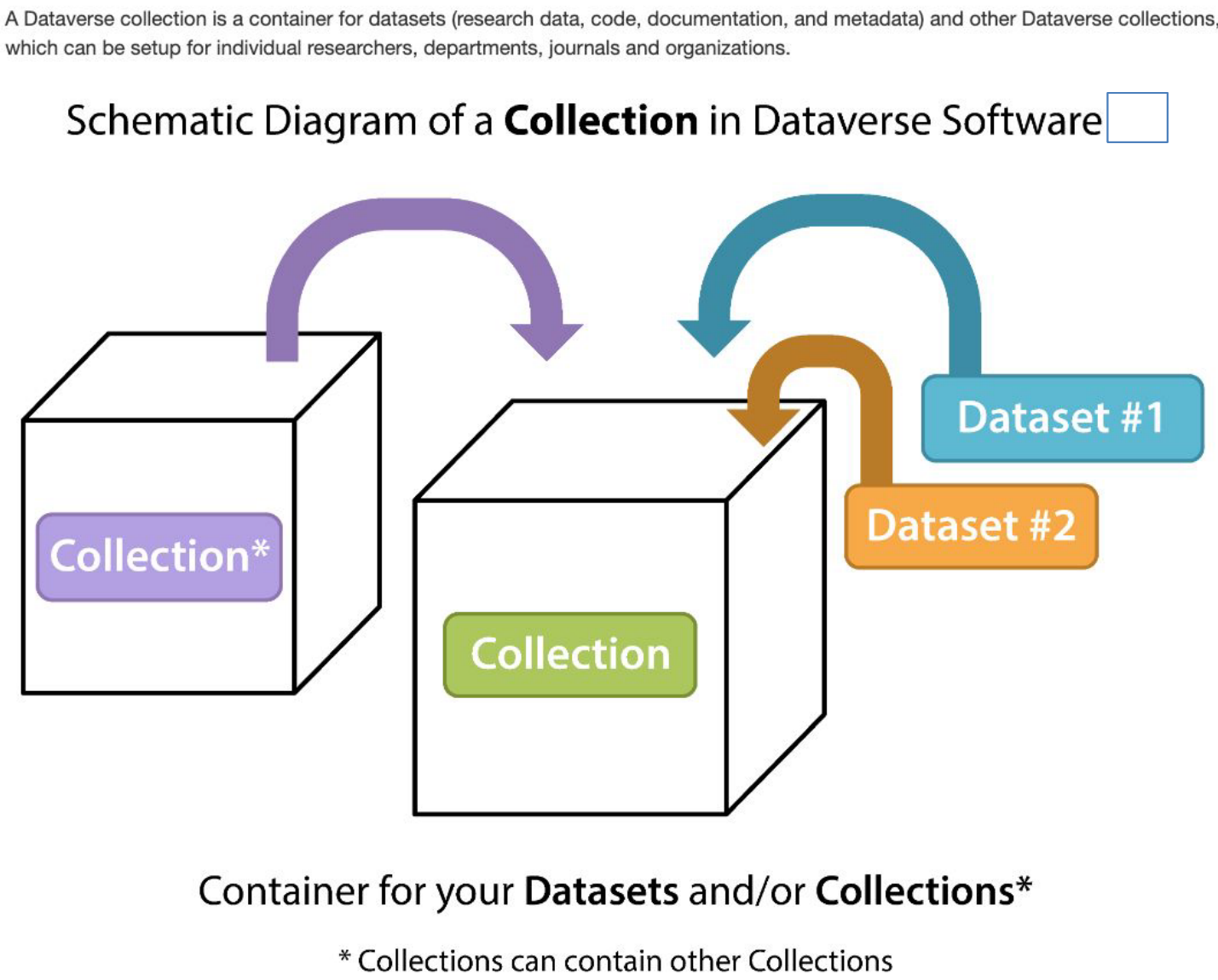
- Support for FAIR Data Principles
- Data citation for datasets and files
- OAI-PMH (Harvesting)
- APIs for interoperability and custom integrations
- API client libraries
- DataCite integration
- Login via Shibboleth
- Login via ORCID, Google, GitHub, or Microsoft
- Login via OpenID Connect (OIDC)
- 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 S3
- 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

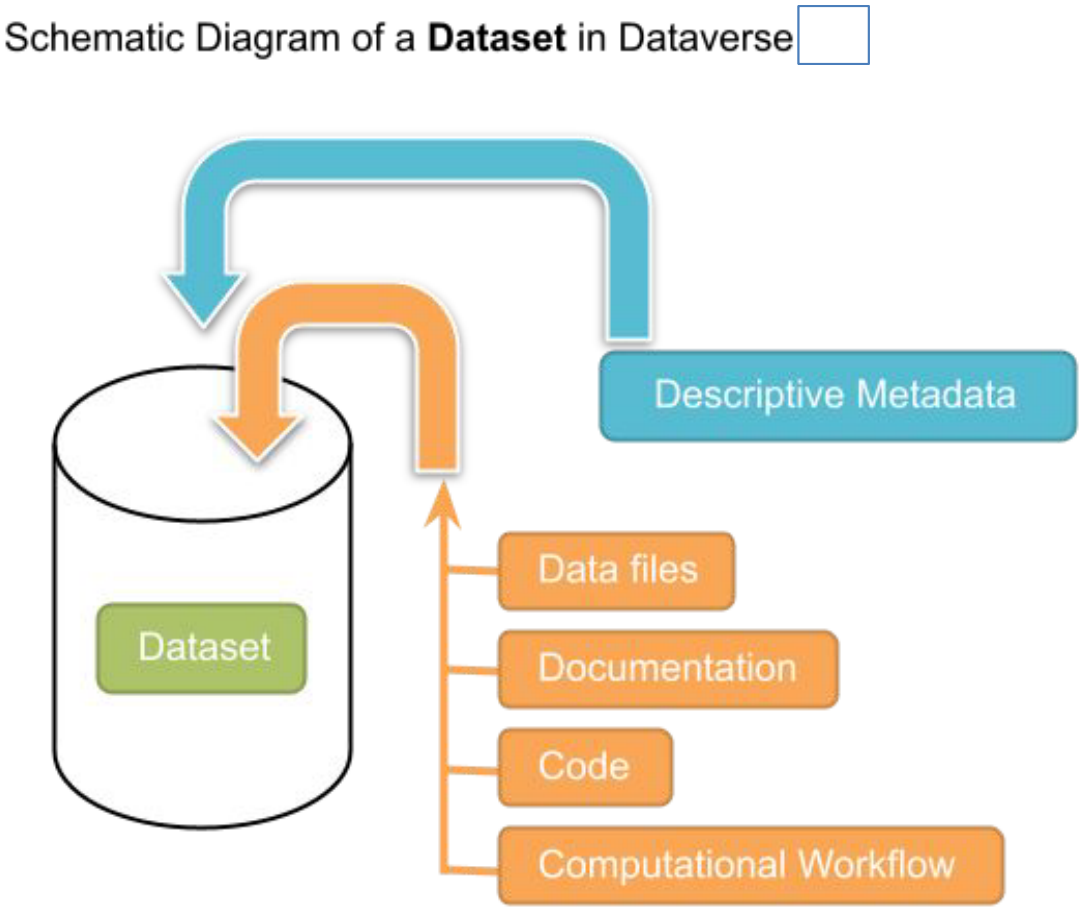
Organizzazione in «collezioni»

Le collezioni di Dataverse contengono i dataset.

Sono personalizzabili con URL, logo, contatti, permessi, set di metadati e filtri di ricerca.



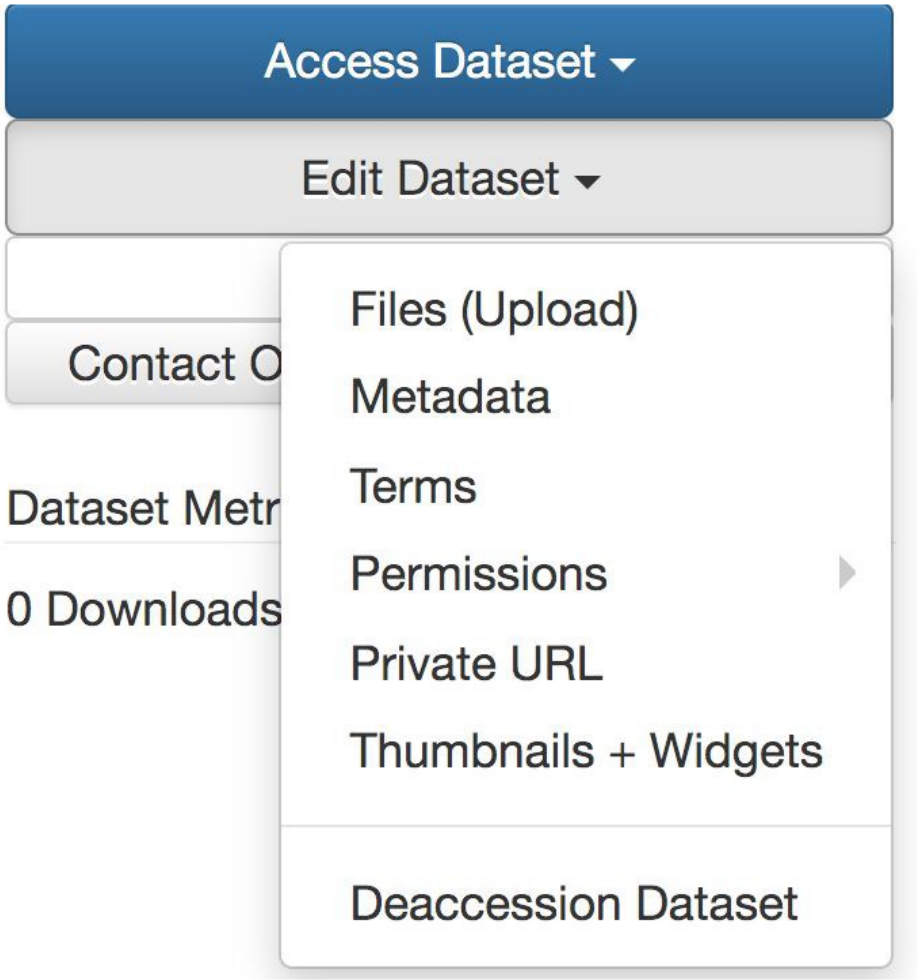
Depositare un dataset



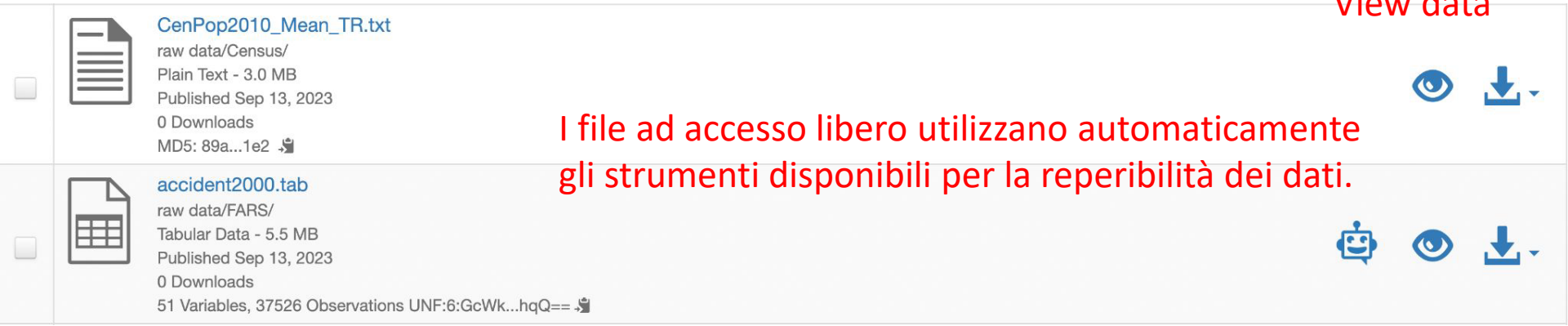
Processo di deposito:
- Scegliere il set di metadati.
- Caricare uno o più file.
- Definire i permessi di accesso.
- Aggiungere thumbnails e widget.
- Inviare a revisione.
- Creare un URL privato per condivisione.
- Definire un embargo.
- Creare versioni.
- Vedere le metriche.
- Pubblicare.
- Ritirare (resta la citazione).

Editing del dataset

testo

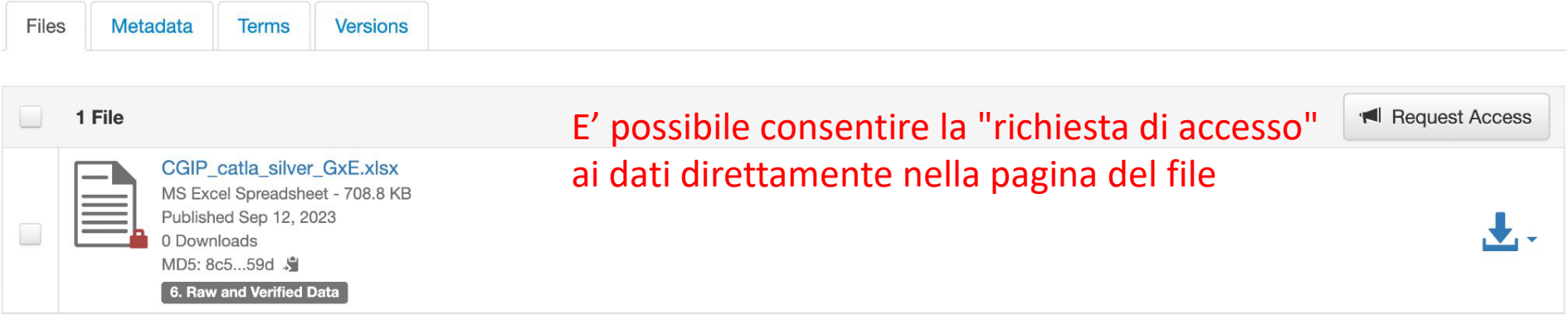


Utilizzo dei dati



I file ad accesso libero utilizzano automaticamente gli strumenti disponibili per la reperibilità dei dati.

Ask the data



E’ possibile consentire la "richiesta di accesso" ai dati direttamente nella pagina del file

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

GeoJSON Previewer

This file is part of "GeoJSON Example".

Version 1.0

File Citation

Durbin, Philip, 2022, "wikipedia.geojson", *GeoJSON Example*, https://doi.org/10.70122/FK2/GFBLSO/S7JA5J, Demo Dataverse, V1

Cite Data File ▾ Learn about Data Citation Standards.

Dataset Citation

Durbin, Philip, 2022, "GeoJSON Example", https://doi.org/10.70122/FK2/GFBLSO, Demo Dataverse, V1

Cite Dataset ▾ Learn about Data Citation Standards.

Access File ▾

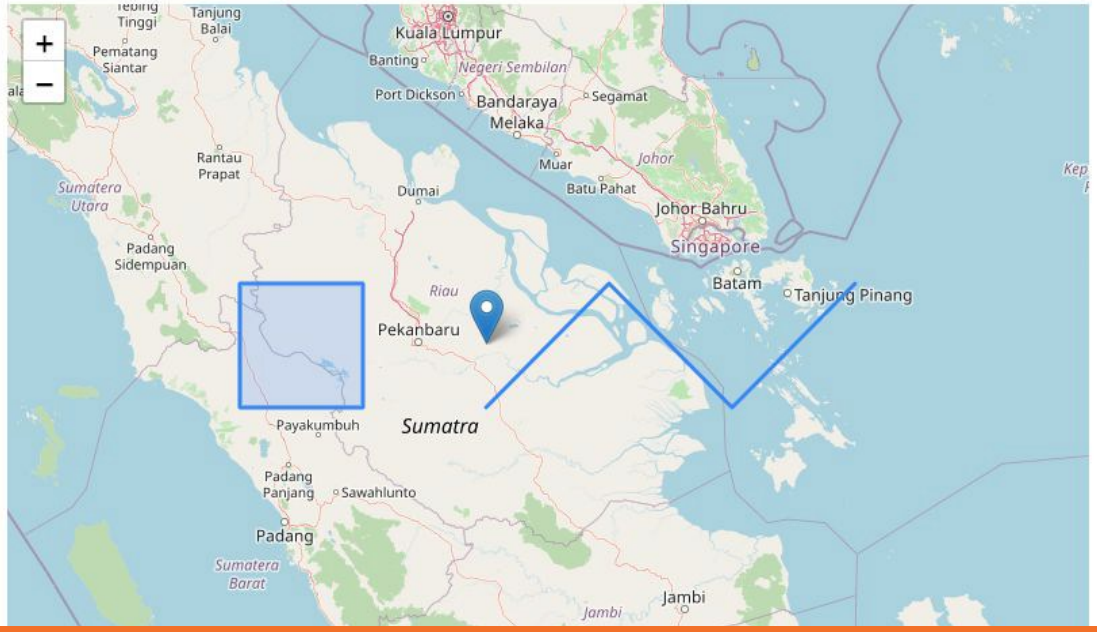
Contact Owner | Share

File Metrics

4 Downloads

Preview | Metadata | Versions

Open in New Window



Tabular Files / Data Explorer

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]

18 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	othereducivl	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

Dataverse e intelligenza artificiale

Le ricerche e lo sviluppo sono iniziati nel 2023 a Harvard e nella comunità Dataverse.
L'accento è posto sull'intelligenza artificiale generativa, LLM, RAG, human-in-the-loop.
Molte risorse, strumenti e progetti sono in fase di produzione o in corso.
Promessa dell'IA: migliorare

- la reperibilità dei dati;
- l'esplorazione e la curatela dei dati- la qualità dei metadati;
- la riutilizzabilità dei dati;
- lo sviluppo e l'integrazione degli strumenti

Photo by Justin Ha on Unsplash

Chi beneficia di Dataverse AI?

Image by Gerd Altmann at Pixabay

Dataverse users
Data depositors
Data reusers
Data seekers
Data curators
Repository managers
Repository strategists
Software developers
System integrators
Metadata harvesters

Risorse AI

Resource, Tool or Project	Type	Improves	Status
AI Guide	Resource	User experience	Production
Ask the Data	Tool	Data reusability	Production
Ask Dataverse	Tool	User experience	Experimental
AutoSage	Tool	Metadata quality, User experience	Experimental
Croissant 🥐	Tool	Data findability	Production
Enhancing Dataset Metadata Project	Project (Tool)	Metadata quality, User experience	Development
GREI AI Taxonomy	Resource	Understanding of AI roles in RDM	Production
Research Data Metadata Knowledge Graph	Project (Tool)	Data findability, Data reusability	Development
Model Context Protocol (MCP) Server	Tool	Data findability, Data reusability	Production
Spam Detection & Workflow Automation	Project (Tool)	Repository data quality	Investigation
TurboCurator	Tool	Metadata quality, User experience	Production

Le sfide attuali e l’evoluzione tecnologica

- Crescita dei volumi di dati
- Diversità dei formati
- Preservazione a lungo termine
- Modelli di sostenibilità

- Verso architetture cloud-native
- API potenziate
- Automazione e metadati machine-actionable

(immagine generata da Gemini)

Innovazioni in corso e prospettive future

- Dataset preview, nuovi metadati, UX migliorata
- Supporto per dati complessi e multimodali

- IA e automazione
- Dataset semantici
- Federazione globale dei Dataverse

(immagine generata da Gemini)

Conclusione: una visione “dall’alto”

Dataverse come infrastruttura abilitante per l’Open Science globale

Connettore tra dati, pubblicazioni e persone

Comunità, collaborazione e apertura come chiavi del futuro

Grazie per l'attenzione!

Susanna.Mornati@4science.com

Harvard Dataverse – https://dataverse.harvard.edu
Community Dataverse – https://dataverse.org
GitHub – https://github.com/IQSS/dataverse

The Dataverse Project

4Science www.4science.com