Generalist Repository Comparison Chart

doi: 10.5281/zenodo.3946720

This chart is designed to assist researchers in finding a generalist repository should no domain repository be available to preserve their research data. Generalist repositories accept data regardless of data type, format, content, or disciplinary focus. For this chart, we included a repository available to all researchers specific to clinical trials (Vivli) to bring awareness to those in this field.

https://fairsharing.org/collection/GeneralRepositoryComparison

TOPIC	HARVARD DATAVERSE	DRYAD	FIGSHARE	MENDELEY DATA	OSF	VIVLI	ZENODO
Brief Description	Harvard Dataverse is a free data repository open to all researchers from any discipline, both inside and outside of the Harvard community, where you can share, archive, cite, access, and explore research data.	Open-source, community-led data curation, publishing, and preservation platform for CC0 publicly available research data Dryad is an independent non-profit that works directly with: • researchers to publish datasets utlizing best practices for discovery and reuse • publishers to support the integration of data availability statements and data citations into their workflows • institutions to enable scalable campus support for research data managment best practices at low cost	A free, open access, data repository where users can make all outputs of their research available in a discoverable, reusable, and citable manner. Users can upload files of any type and are able to share diverse research products including datasets, code, multimedia files, workflows, posters, presentations, and more. With discoverable metadata supporting FAIR principles, file visualizations, and integrations, researchers can make their work more impactful and move research further faster.	Mendeley Data is a free repository specialized for research data. Search more than 20+ million datasets indexed from 1000s of data repositories and collect and share datasets with the research community following the FAIR data principles.	OSF is a free and open source project management tool that supports researchers throughout their entire project lifecycle in open science best practices.	Vivli is an independent, non-profit organization that has developed a global data-sharing and analytics platform. Our focus is on sharing individual participant-level data from completed clinical trials to serve the international research community.	Powering Open Science, built on Open Source. Built by reserachers for researchers. Run from the CERN data centre, whose purpose is long term preservation for the High Energy Physics discipline, one of the largest scientific datasets in the world
Size limits	No byte size limit per dataset. Harvard Dataverse currently sets a file size limit of 2.5GB.	300GB/dataset	Soft limit of 20GB/file for free accounts. System limit of 5000GB/file. Unlimited storage of public data but 20GB storage for private data for free accounts. Email info@figshare.com to have upload and storage limits raised.	10GB per dataset	Projects currently have not storage limit. There is a 5GB/file upload limit for native OSF Storage. There is no limit imposed by OSF for the amount of storage used across add-ons connected to a given project.	If more than 10GB per study data, reach out to us	50GB per dataset, contact us via https://zenodo.org/support for higher limits
Storage space per researcher	1 TB per researcher	No limit	No limit	No limit	No limit	No limit	No limit
Persistent, Unique Identifier Support	DOI, Handle	DOI	DOI	DOI	DOI	DOI	DOI

TOPIC	HARVARD DATAVERSE	DRYAD	FIGSHARE	MENDELEY DATA	OSF	VIVLI	ZENODO
Licensing Options	By default, datasets are published in the public domain (CC0). Depositors can change this to apply their own licenses.	CC0	Default licenses supported: CC0 1.0 CC BY 4.0 MIT Apache 2.0 GPL v3 GPL v2	Default licenses supported: CC0 1.0 CC BY 4.0 CC BY NC 3.0 MIT Apache 2.0 BSD 3-Clause BSD 2-Clause GPL v3 GPL v2 LGPL MPL-2.0 CeCILL CeCILL-B CERN OHL TAPR OHL	The following 14 licenses are available: No License - is a copyright license for the project authors and contributors CC0 1.0 CC-By 4.0 MIT Apache 2.0 BSD 2-Clause BSD 3-Clause GPL 3.0 GPL 2.0 Artistic 2.0 Eclipse 1.0 LGPL 3.0 LGPL 2.1 Mozilla 2.0 Other- user defines a license in a .txt file and uploads to the project (not available on registrations or collections)	https://vivli.org/resources/vivli-data-use-agreement/	Content is available publicly under any one of 400 open licences (from opendefinition.org and spdx.org). Restricted and Closed content is also supported.
Costs to the researcher	Harvard Dataverse is free for all researchers worldwide (up to 1 TB)	Costs covered by institutional, publisher, and funder members, otherwise a one-time fee of $120 for authors to cover cost of curation and preservation	Free	Free	Free	https://vivli.org/about/share-data/ All fees to share, store, and access COVID-19 clinical research are waived.	Free up to 50GB per dataset. Larger datasets possible by arrangement

TOPIC	HARVARD DATAVERSE	DRYAD	FIGSHARE	MENDELEY DATA	OSF	VIVLI	ZENODO
Equitable free and ongoing access to data	Use of CC0 and free data access highly encouraged. Support for broad indexing and long-term preservation strategies	CC0, public available, broadly indexed research data with long term preservation in a CoreTrustSeal-certified repository	Metadata is CC0. All files and metadata can be accessed from docs.figshare.com. Figshare has also partnered with DuraSpace and Chronopolis to offer further assurances that public data will be archived under the stewardship of Chronopolis. In the highly unlikely event of multiple AWS S3 failures, Figshare can restore public user content from Chronopolis.	Metadata is licensed CC0. Datasets are and will continue to be free access Long-term access in the event of cease of operations granted by DANS. Access to archived datasets will be provided for free in perpetuity	Open, public API to support broad indexing, partnership with Internet Archive for long-term preservation, $250k preservation fund.	Managed Access as Human Subject Data. No charge period for data that is accessible only with a research environment. Costs after no charge time period ends. https://vivli.org/resources/vivli-secure-research-environment-2/	Metadata is licensed CC0. Content is both online on disk and offline on tape as part of long-term preservation policy
Version Support	Yes, including version comparison and W3C provenance support	Yes, versioning is supported	Yes	Yes, including version comparison, to easily see what has changed from version to version	Yes for OSF Storage, when supported for integrated storage providers	Coming soon	Yes, with "Concept" DOI to represent "all versions"
Characteristics - Standards							
Supported Metadata Schemas	Dublin Core, Data Documentation Initiative (DDI Codebook), DataCite, OpenAIRE, Schema.org, OAI_ORE	DataCite, schema.org	Dublin Core (oai_dc), Datacite (oai_datacite), RDF (rdf), CERIF XML (cerif), Qualified Dublin Core (qdc) (hasPart support), Metadata Encoding and Transmission Standard (mets) and UKETD_DC (uketd_dc)	Dublin Core, DataCite, Schema.org	Datacite, Crossref preprint	DataCite	DataCite
Formats Supported for Export	JSON, XML	Fully documented API available for direct integration. Exports available in JSON, XML, schema.org	Dublin Core (oai_dc), Datacite (oai_datacite), RDF (rdf), CERIF XML (cerif), Qualified Dublin Core (qdc) (hasPart support), Metadata Encoding and Transmission Standard (mets) and UKETD_DC (uketd_dc)	JSON, XML	JSON, API, JSONAPI		DataCite, Dublin Core, DCAT-AP, JSON, JSON-LD, GeoJSON, MARCXML, Citation Style Language JSON + support for custom metadata formats

TOPIC	HARVARD DATAVERSE	DRYAD	FIGSHARE	MENDELEY DATA	OSF	VIVLI	ZENODO
Supported Community Vocabulary or Taxonomy	ISO 3166-1 CV for geospatial metadata, ISO 639-1 CV for languages, DataCite's dataset contributor vocab, and subsets of the OBI Ontology and NCBI Taxonomy for Organisms	ORCID - Authors ROR - Institutions Open Funder Registry - Funding PLOS Thesaurus - Keywords	ORCID - Authors, Dimensions - Funding, Field of Research codes, GRID ids, Implementing ROR and Make Data Count	Omniscience taxonomy for subject categories and keywords, and additional custom vocabularies can be loaded on institutional version	BePress 3 tier taxonomy for preprints and custom taxonomy mapped to Bepress on preprints	Cochrane Linked Data Vocabulary	ORCID - Authors, FundRef + OpenAIRE Projects Database - Funding
Characteristics Useful for Linking data to other relevant digital information							
Support for creators/ authors identifiers	ORCID, ISNI, LCNA, VIAF, GND, DAI, ResearcherID and ScopusID	ORCID required for corresponding author, co-author ORCID supported	ORCID	ORCID - Mendeley ID - Scopus Author ID	ORCID, ResearcherID	ORCID	ORCID
Support linking to related publications	Yes	DataCite relation types	DataCite relation types	Yes, DataCite and Scholix relation types	Preprints link to related Peer-reviewed publications using Crossef 'isPrerint of'	Yes	DataCite relation types
Linking of derived products from another	Yes	DataCite relation types	DataCite relation types	Yes, DataCite and Scholix relation types	No		DataCite relation types
Grant ID(s)	Yes	Yes	Yes, hard-linked to Dimensions.ai database	Yes, customizable on institutional version	No		Yes
Grant ID affiliation(s)	Yes	Open Funder Registry	Yes, hard-linked to Dimensions.ai database	Yes, customizable on institutional version	No		Yes
Creator/author affiliations(s)	Yes	ROR for all authors	GRID, implementing ROR	Yes, Mendeley Institution IDs (Ringgold)	No	Yes	Yes
Linking to related software	Yes	DataCite relation types	DataCite relation types	Yes with semantic links	No		DataCite relation types
Linking to related research product	Yes	DataCite relation types	DataCite relation types	Yes with semantic links	No		DataCite relation types
Characteristics supporting Metrics							
Supported Data Use Metrics	Downloads, explorations, pageviews, data volume	Investigations (Views), Requests (Downloads), citations	View, downloads, citations, altmetrics	Views, downloads, altmetrics	Downloads (per version), Links, Forks	https://vivli.org/resources/platform_metrics-2/	Views, Downloads, Data Volume, Citations, Altmetrics

TOPIC	HARVARD DATAVERSE	DRYAD	FIGSHARE	MENDELEY DATA	OSF	VIVLI	ZENODO
Normalized data usage statistics (e.g. Make Data Count)	By end of Q2, will enable Counter Code of Practice for Reseach Data (Make Data Count)	Following standards: Counter Code of Practice for Reseach Data (Make Data Count) in both standardizing and reporting usage to DataCite	Currently enabling Counter Code of Practice for Reseach Data (Make Data Count)	Following standards: Counter Code Of Practice for Research Data Usage Metrics, starting Q3 2020 (Make Data Count)	No		Following standards: Counter Code Of Practice for Research Data Usage Metrics (Make Data Count)
Characteristics supporting protection and related journal articles							
Support embargo	Yes, through draft versions	So long as the related journal requires or allows this	Yes	Yes	Yes for registrations only	Yes	Yes
Supports peer review during embargo (e.g. "temporary share link")	Yes	Yes, Private for peer review URL available	Yes	Yes	View only link with ability to anonymize contributor list	Yes	Yes
Support managed access	Yes, with request access workflow	Only for peer review pre-publication	Yes	Yes	Yes with request access and private sharing setting	Yes	Yes, with access request workflow
Preservation	Harvard Dataverse preserves files at Harvard FAS Research Computing), Reformat upon ingest to preservation formats (for some proprietary original formats)	Core Trust Seal Certified Merritt repository with storage in US and EU: at San Diego Supercomputing Center, DANS, and Zenodo	Chronopolis	DANS. Storage in perpetuity	$250k Preservation fund, IMLS grant for transfer to Internet Archive (In progress)	Vivli has funds available and the assurances of Microsoft that it will maintain and archive data for the lifetime of its use, up to 20 years.	CERN Tape Archive
Human subject research data at the individual participant level	No support for preserving/publishing data with identifying human information	No support for identifying human information	No support for identifying human information	No support for identifying human information	No	Yes, anonymized individual-level patient data in clinical research made available to qualified researchers.	Any anonymous or anonimised data

TOPIC	HARVARD DATAVERSE	DRYAD	FIGSHARE	MENDELEY DATA	OSF	VIVLI	ZENODO
Other Information							
Business Model	Support from Harvard University, Public and Private Grants, and an emergent Consortium model	Non-profit with community memberships (institutions, publishers, funders), direct funder support through grants, and data publishing charges	Institutional, Funder, Governmental service subscriptions	Subscription model for Academic & Government entities	Non-profit with direct funder support through grants, government contracts, and community memberships.	Funded via grants and member fees.	Base infrastructure by CERN, a non-profit IGO. Projects through Grants
Links to Registries for more information							
FAIRsharing record	https://doi.org/10.25504/FAIRsharing.t2e1ss	https://doi.org/10.25504/FAIRsharing.wkggtx	https://doi.org/10.25504/FAIRsharing.drtwnh	https://doi.org/10.25504/FAIRsharing.3epmpp	https://doi.org/10.25504/FAIRsharing.g4z879	https://doi.org/10.25504/FAIRsharing.uovQrT	https://doi.org/10.25504/FAIRsharing.wy4egf
re3data record	http://doi.org/10.17616/R3C880	http://doi.org/10.17616/R34S33	http://doi.org/10.17616/R3PK5R	http://doi.org/10.17616/R3DD11	http://doi.org/10.17616/R3N03T	http://doi.org/10.17616/R3SB9S	http://doi.org/10.17616/R3QP53