
Introducción a Dataverse

Gustavo Durand

Technical Lead / Architect

IQSS, Harvard University



HARVARD
LIBRARY



HARVARD UNIVERSITY
Information Technology

Introducción a Dataverse

Descripción

- Una plataforma de código abierto para publicar, citar y archivar datos de investigación
- Creada para soportar múltiples tipos de datos, usuarios y flujos de trabajo
- Desarrollada en Harvard's Institute for Quantitative Social Science (IQSS) desde 2006
- IQSS financia su desarrollo y fondos de proyectos, en colaboración con instituciones alrededor del mundo
- Equipo central @ IQSS - desarrolladores, diseñadores, UX/UI, especialistas en metadatos , equipo de curación, equipo de liderazgo

Dataverse: Características

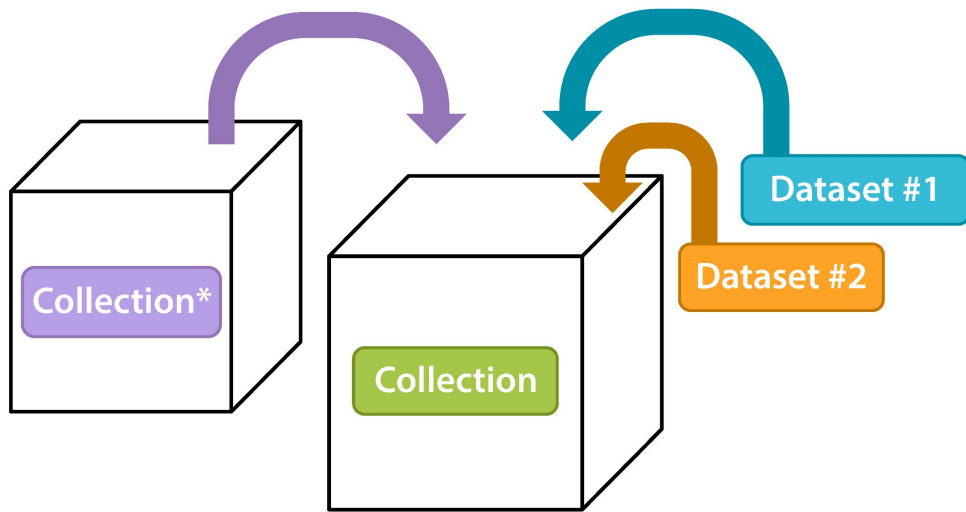
<https://dataverse.org/software-features>

- El objetivo central del código es focalizarse en publicación (citación, compartición, versionado, etc.), Principios Datos FAIR
- APIs robustas que permiten interoperabilidad con “herramientas externas” y otros repositorios / software

Colecciones Dataverse

- Habilidad de crear colecciones en Dataverse para organizar conjuntos de datos (datasets) de acuerdo a las necesidades
- Las colecciones Dataverse pueden contener otras colecciones permitiendo estructuras jerárquicas
- Se pueden aplicar reglas diferentes a distintas colecciones en Dataverse, por ej. Metadatos, Permisos, etc.

Schematic Diagram of a **Collection** in Dataverse Software 5.0



Container for your **Datasets** and/or **Collections***

* Collections can contain other Collections

Metadatos Dinámicos

- Los metadatos se definen dinámicamente a nivel de base de datos, permitiendo agregar nuevos bloques de metadatos de forma modular.
- Permite:
 - Valores individuales o múltiples
 - Valores simples o compuesto
 - Vocabularios controlados

Choose the metadata fields to use in dataset templates and when adding a dataset to this dataverse.

- ☒ Citation Metadata (Required) [\[+\] View fields + set as hidden, required, or optional](#)
- ☐ 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](#)

Citation Metadata ^

Title * ⓘ

Author * ⓘ

Name * ⓘ Affiliation ⓘ

Identifier Scheme ⓘ Identifier ⓘ

Contact * ⓘ

Name ⓘ Affiliation ⓘ

E-mail * ⓘ

Description * ⓘ

This field supports only certain [HTML tags](#).

Text * ⓘ

Date ⓘ

Sistema de Permisos Flexibles

- Permite múltiples flujos de trabajo controlando quiénes pueden agregar a una colección Dataverse, qué pueden hacer y qué rol tienen en los datasets.
- Los roles son definidos como una serie de permisos para usuarios o grupos
- Los grupos se pueden definir de forma estática o dinámica (usuarios que se loguean desde la misma institución, via Shibboleth)

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?

- ☒ Contributor - 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

2 Users/Groups

User/Group Name (Affiliation) ⇅	ID ⇅	Role ⇅
Dataverse Admin (Dataverse.org)	@dataverseAdmin	Admin
Anyone with a Dataverse account	:authenticated-users	Dataverse + Dataset Creator

APIs Robustas

- APIs para búsqueda, depósito, acceso, administración, métricas, etc.
- APIs adicionales para cosecha (descubrimiento) e interoperabilidad con otros sistemas
- Herramientas externas se pueden registrar por medio de APIs, de modo que Dataverse puede generar enlaces en la UI, así los usuarios pueden previsualizar, explorar, configurar y más

API Guide

Contents:

- [Introduction](#)
 - [What is an API?](#)
 - [Types of Dataverse Software API Users](#)
 - [API Users Within a Single Dataverse Installation](#)
 - [Users of Integrations and Apps](#)
 - [Power Users](#)
 - [Support Teams and Superusers](#)
 - [Sysadmins](#)
 - [In House Developers](#)
 - [API Users Across the Dataverse Project](#)
 - [Developers of Integrations, External Tools, and Apps](#)
 - [Developers of Dataverse Software API Client Libraries](#)
 - [Developers of The Dataverse Software Itself](#)
 - [How This Guide is Organized](#)
 - [Getting Started](#)
 - [API Tokens and Authentication](#)
 - [Lists of Dataverse APIs](#)
 - [Client Libraries](#)
 - [Examples](#)
 - [Frequently Asked Questions](#)
 - [Getting Help](#)
- [Getting Started with APIs](#)
 - [Servers You Can Test With](#)
 - [Getting an API Token](#)
 - [curl Examples and Environment Variables](#)
 - [Depositing Data](#)
 - [Creating a Dataverse Collection](#)
 - [Creating a Dataset](#)
 - [Uploading Files](#)
 - [Publishing a Dataverse Collection](#)
 - [Publishing a Dataset](#)

Tecnología Dataverse

Payara 5*



Java 11

Java EE8*

- Presentación: JSF (PrimeFaces), RESTful API
- Business: EJB, Transactions, Asynchronous, Timers
- Almacenamiento: JPA (Entities), Bean Validation

Almacenamiento: Postgres, Solr, File System / Swift / S3

Comunidad Dataverse

Comunidad Dataverse

- 140+ Contribuyentes en Github
- Cientos de miembros de la Comunidad Dataverse - desarrolladores, investigadores, bibliotecarios, cientistas de datos
 - Workshops & Entrenamientos
 - UX/UI Testeo & Entrevistas
 - Global Dataverse Community Consortium
 - Dataverse Google Group / Matrix / Community Slack
 - Llamadas de la Comunidad Dataverse
 - Reuniones de la Comunidad Dataverse

Global Dataverse Community Consortium

- Apoya a los repositorios Dataverse en todo el mundo
- El Global Dataverse Community Consortium (GDCC) se dedica a brindar una organización internacional a los esfuerzos existentes de la comunidad Dataverse, y proporcionará un lugar de colaboración para que las instituciones aprovechen las economías de escala para apoyar a los repositorios de Dataverse en todo el mundo.



<http://DataverseCommunity.Global>

Comunidad Dataverse

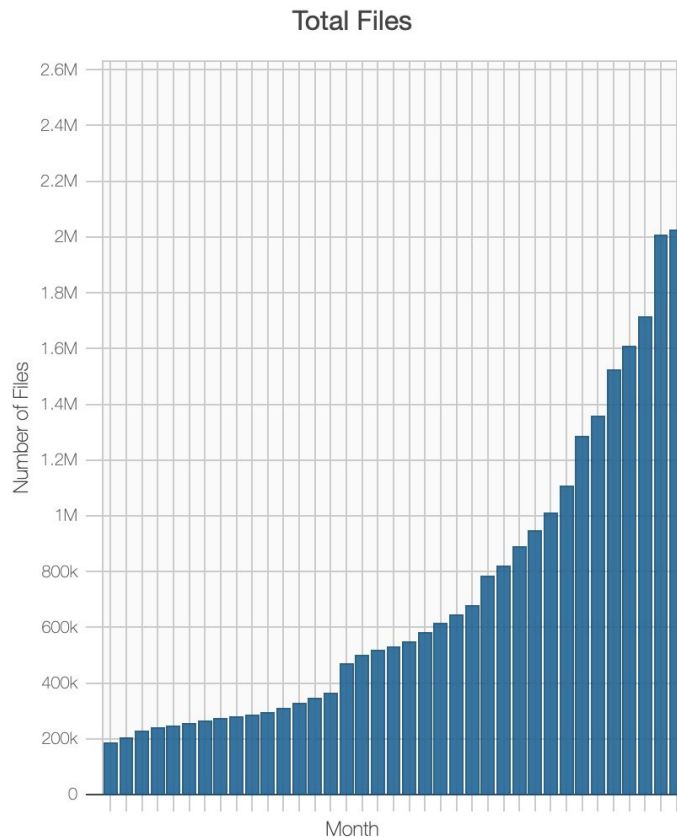
- 83 instalaciones informadas en todo el mundo



Los Datos (dataverse.org/metrics)

- 83 instalaciones
- 13.000 Colecciones Dataverse*
- 210.000 Datasets*
- 2,290.000 Archivos*
- 68.300.000 Descargas de Archivos*

* métricas recopiladas de 61 instalaciones
(ejecutando 4.9 y más reciente)



Planes Futuros

Actividades para el equipo Dataverse @ Harvard

- Harvard Data Commons (HDC)
- NIH GREI Program Activities
 - Propuesta individual
 - Actividades “Coopetición”
 - Dataverse, Dryad, Figshare, Mendeley Data, Open Science Framework, Vivli
- Soporte de Harvard Dataverse
- Facilitar el Desarrollo de Comunidades

Harvard Data Commons

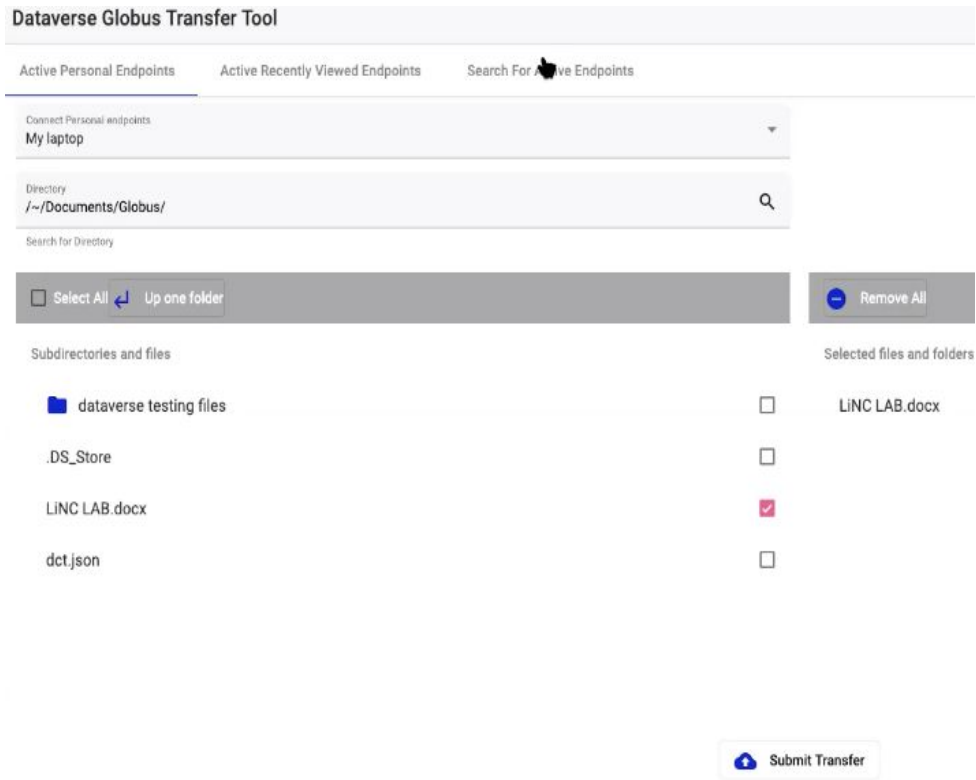
Automatización del flujo de datos de investigación desde entornos informáticos de investigación hacia entornos de gestión, publicación, descubrimiento y preservación

Tres objetivos principales:

- Automatización de la conexión técnica entre las infraestructuras informáticas de investigación y Dataverse
- Mejora de Dataverse para admitir flujos de trabajo accionables por máquina de varios tipos
- Automatizar las conexiones entre los sistemas de investigación y los sistemas bibliotecarios clave utilizados para archivo y la publicación

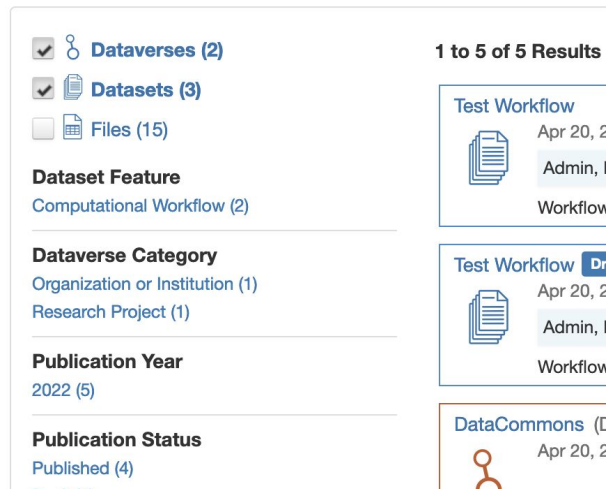
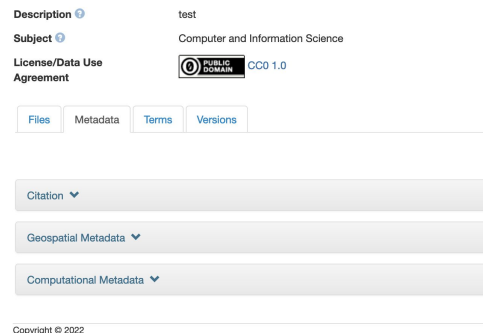
Integración con Globus

- Habilidad de agregar archivos (o solo metadatos de archivos, dejando los archivos en la fuente) a Dataverse
- Desarrollada como una herramienta externa a integrarse en el flujo de trabajo de carga de Dataverse



Soporte de flujo de trabajo computacional

- Existen herramientas externas para respaldar los flujos de trabajo y la reproducibilidad
- Se agregaron metadatos adicionales a Dataverse para mejorar la capacidad de descubrimiento (por ejemplo, la nueva faceta "Característica de conjunto de datos")
- Validación automática de checksum en la subida de archivos BagIt (basado en el manifiesto BagIt)



Mejora de la conexión entre Sistemas

- Nuevas funciones de archivo más sólidas para S3
 - Incluye nueva API de estado de archivo y pantalla de administración
- Notificación bidireccional de recursos relacionados
 - Capacidad para enviar y recibir notificaciones de datos vinculados (linked data) sobre las relaciones entre conjuntos de datos <--> artículos/otros recursos
 - Eventualmente cumplirá con el protocolo COAR Notify
 - Característica adicional: capacidad de agregar instrucciones personalizadas a las plantillas

NIH GREI Año 1 Actividades del programa

- Soporte remoto de Large Storage
- Vocabularios Controlados para Biomédicas
- Descubrimiento para DDI-CDI
- Software y flujos de trabajo biomédicos
- Cosecha y compartición de metadatos entre repositorios
- Métricas de Uso - Soporte para Make Data Count
- Revisión de soporte de datos confidenciales
- Evaluación y Evolución de la Arquitectura
- Planes de Gestión de Datos para NIH
- Capacitación y divulgación (por ejemplo, ¡esta reunión!)

Hoja de Ruta de Dataverse

<https://www.iq.harvard.edu/roadmap-dataverse-project>

- Metas estratégicas
- Implementación, Planificación, Futuro

Muchas gracias



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 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?](#)



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 Dataverse as part of your journal data policy or list of repository recommendations. [Want to find out more about journal dataverses?](#)



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 Dataverse. [Want to contribute?](#)

<https://dataverse.org>
<https://github.com/iqss/dataverse>