
Dataverse (y Los PIDs)

Gustavo Durand

Líder técnico / Arquitecto

IQSS, Harvard University

<https://doi.org/10.5281/zenodo.7860265>

#PIDsLATAM23



Dataverse (y Los PIDs)

Gustavo Durand

Líder técnico / Arquitecto

IQSS, Harvard University



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
 - Contribuidores claves de la comunidad

Dataverse: Características

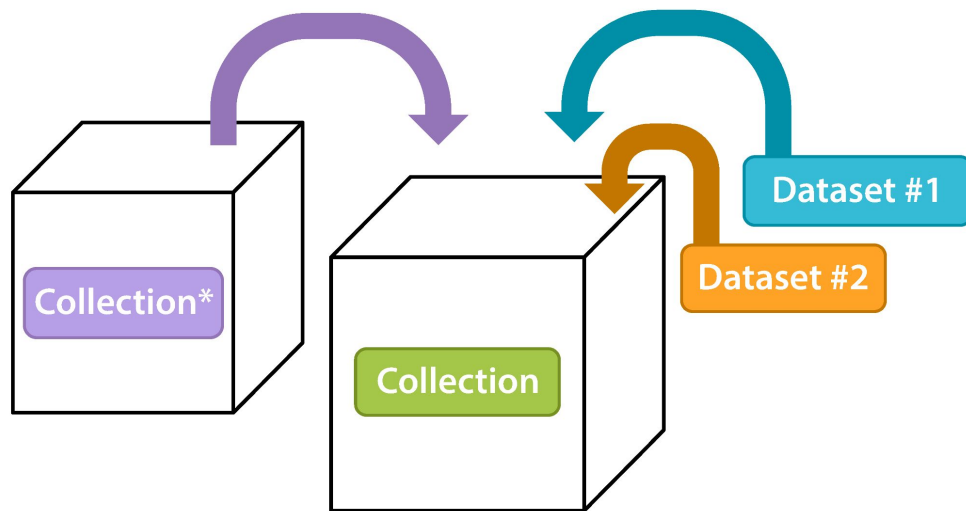
<https://dataverse.org/software-features>

- 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

- 98 instalaciones informadas en todo el mundo



Comunidad Dataverse

- 160+ Contribuyentes en Github
- Cientos de miembros de la Comunidad Dataverse - desarrolladores, investigadores, bibliotecarios, científicos 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
 - La Copa de Dataverse 

La Copa de 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>

Los Datos (dataverse.org/metrics)

- 98 instalaciones
- 13.900 Colecciones Dataverse*
- 204.000 Datasets*
- 2,640.000 Archivos*
- 72.400.000 Descargas de Archivos*

* métricas recopiladas de 59 instalaciones
(ejecutando 4.9 y más reciente)

Datasets by Most Common Subject



Identificadores Persistentes en Dataverse

5 PIDs básicos para Potenciar la Visibilidad

5 Core PIDs to Power Findability

1. Digital object identifiers (DOIs)
2. Open Researcher and Contributor (ORCID) iDs
3. Research Organization Registry (ROR) IDs
4. Crossref Funder Registry IDs
5. Crossref Grant IDs

- Del artículo “Implementing Effective Data Practices: Stakeholder Recommendations for Collaborative Research Support”
- <https://www.arl.org/wp-content/uploads/2020/09/2020.09.25-implementing-effective-data-practices.pdf>

Identificadores digitales de Objetos (DOIs)

- **Digital object identifiers (DOIs)** para identificar datos, así como publicaciones y otros resultados
 - DOIs para Datasets via DataCite
 - DOIs para Archivos (configurable en la instalación)
- También ofrece soporte para otros proveedores PID para datos, ej. handles
- Planes y intereses futuros:
 - Configuración de DOIs de Archivos por colección Dataverse
 - DOIs para Versiones de Dataset
 - Diferentes prefijos por colección Dataverse



Müller, Martin; Gogishvili, David; Wolfe, Sven Daniel, 2022, "Dataset: Revenues and Costs of the Olympic Games and the Football World Cups, 1960s to late 2010s", <https://doi.org/10.7910/DVN/CPQEHN>, Harvard Dataverse, V2

Cite Dataset ▾

Learn about [Data Citation Standards](#).

1 Work

Dataset: Revenues and Costs of the Olympic Games and the Football World Cups, 1960s to late 2010s

Martin Müller, David Gogishvili & Sven Daniel Wolfe

Content published 2022 in [Harvard Dataverse](#)

Dataset on major revenues (ticketing, sponsoring, broadcasting) and costs (organisation, sports venues) for the Olympic Games (Summer and Winter) and the Men's Football (Soccer) World Cup from the 1960s to the late 2010s. Data are in original currency and in USD2018 reported. Sources are reported.

DOI registered May 9, 2022 via DataCite.

Open Researcher and Contributor (ORCID) iD

- **Open Researcher and Contributor (ORCID) iDs** para identificar investigadores
- Autenticación
 - Permite al usuario loguearse via su cuenta ORCID
 - Guardar ORCID iD con la cuenta
- Planes y intereses futuros
 - Permitir a los usuarios pre cargar el dataset con su ORCID iD
 - Permitir a los usuarios buscar autores/ contribuidores por ORCID iD

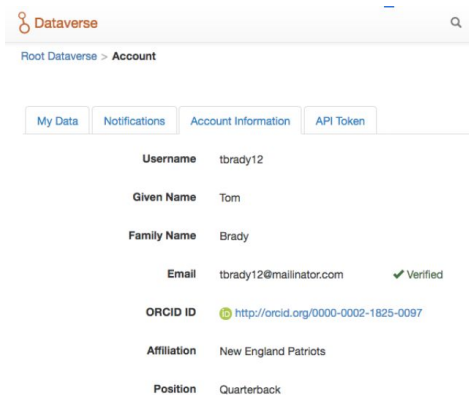
➔ Log In

Log in or sign up with your ORCID account — [more information about account creation](#). Having trouble? Please contact [DataverseNO Support](#) for assistance.


 Create or Connect your ORCID

ORCID is an open, non-profit, community-based effort to provide a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers. ORCID is unique in its ability to reach across disciplines, research sectors, and national boundaries and its cooperation with other identifier systems. Find out more at orcid.org/about.

This repository uses your ORCID for authentication (so you don't need another username/password combination). Having your ORCID associated with your datasets also makes it easier for people to find the datasets you have published.



The screenshot shows the Dataverse user interface. At the top, there is a navigation bar with the Dataverse logo and a search icon. Below the navigation bar, the breadcrumb trail reads "Root Dataverse > Account". There are four tabs: "My Data", "Notifications", "Account Information", and "API Token". The "Account Information" tab is active, displaying the following user details:

Username	tbrady12
Given Name	Tom
Family Name	Brady
Email	tbrady12@mailinator.com ✓ Verified
ORCID ID	 http://orcid.org/0000-0002-1825-0097
Affiliation	New England Patriots
Position	Quarterback

Sign in

Email or 16-digit ORCID ID

example@email.com or 0000-0001-2345-6789

Password

SIGN IN

Forgot your password or ORCID ID?

Don't have an ORCID iD yet? [Register now](#)

Soporte de ROR y Crossref

- Soporte a otros PIDs a través de funcionalidades de Vocabularios Controlados Externos
 - **Research Organization Registry (ROR) IDs** para identificar afiliaciones de organizaciones de investigación
 - **Crossref Funder Registry IDs** para identificar financiadores de investigación
 - **Crossref Grant IDs** para identificar becas y otros tipos de premios a investigaciones

Affiliation

Select a research organization ▲

Harvard University

Harvard University, 03vek6s52,

Harvard University Press, 006v7bf86,
HUP

IIT@Harvard, 044hpwe09,

Arnold Arboretum, 05xsxgs79,

Harvard NeuroDiscovery Center

 <https://ror.org/03vek6s52>

Harvard University

ORGANIZATION TYPE

Education

OTHER NAMES

Universidad de Harvard

WEBSITE

<http://www.harvard.edu/>

LOCATION

Cambridge (GeoNames ID [4931972](#))
United States

OTHER IDENTIFIERS

GRID [grid.38142.3c](#)
ISNI [0000 0004 1936 754X](#)
Crossref Funder ID [100007229](#)
Wikidata [Q13371](#)

Planes Actuales y Futuros

Actividades para el equipo Dataverse @ Harvard

- Soporte de Harvard Dataverse
- Facilitar el Desarrollo de Comunidades
- NIH GREI Actividades del programa
 - Propuesta individual
 - Actividades “Coopetición”
 - Dataverse, Dryad, Figshare, Mendeley Data, Open Science Framework, Vivli
- Otros proyectos y colaboraciones relacionadas con Dataverse

NIH GREI 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!)

Planes de nuevas versiones

- En **Mayo / Junio**, estamos planeando lanzar a la vez
 - **5.14**
 - Consistente en nuevas características, solución de bug y contribuciones de la comunidad
 - **6.0**
 - Funcionalmente igual a 5.14, pero corre en el último servidor Payara 6 y otras nuevas tecnologías
 - **Más allá de 6.0**
 - Proyecto de re arquitectura
 - Separar front end en diferentes SPA
 - Expandir el back end de Dataverse y APIs
 - Expandir Modularidad para potenciar a la comunidad

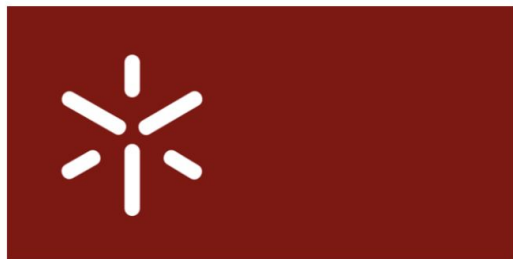
Hoja de Ruta de Dataverse

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

- Metas estratégicas
- Implementación, Planificación, Futuro

Muchas gracias

Reunión de la Comunidad de Dataverse 2023
5-7 de junio, 2023
Universidade do Minho en Braga, Portugal



Universidade do Minho

<https://projects.iq.harvard.edu/dcm2023>



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>