

Sharing Data with The Harvard Dataverse Repository

Danny Brooke,
Program Manager, Product Development

Sonia Barbosa
Manager of Data Curation, Harvard Dataverse Repository
Manager of the Murray Research Archive

Institute for Quantitative Social Science, Harvard University




Agenda

- Dataverse Project Background
- Dataverse Project APIs
- Demo
- Questions

enabling better, bigger, faster,
more collaborative science

enabling better, bigger, faster,
more collaborative science

Dataverse: Open Source Research Data Repository Software

 Dataverse Project


About▼

Community▼

Best Practices▼


Software▼

Contact




The
Dataverse[®]
Project

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?](#)

Dataverse Installations Worldwide

67 Installations



Depositing your Data Into the Harvard Dataverse Repository (in a few easy steps!)


- Create an Account (in the UI)
- Create an API Token (in the UI)
- Create a Dataverse collection (optional, but recommended!)
- Create a Dataset
- Add File(s)
- Publish your Dataset (and Dataverse collection)

Step 1: Create an Account

Account Information

You can also create a Dataverse account with one of our other [log in options](#).

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 ("Agreement").

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

Create Account

[Cancel](#)

Log In

Log in or sign up with your institutional account — more [information about account creation](#). 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

GitHub

Google

ORCID

[Sign up for a Dataverse account.](#)

Step 2: Create an API Token

My Data

Notifications

Account Information

API Token

Your API Token is valid for a year. Check out our [API Guide](#) for more information on using your API Token with the Dataverse APIs.

API Token for Daniel Brooke has not been created.

Create Token

My Data

Notifications

Account Information

API Token

Your API Token is valid for a year. Check out our [API Guide](#) for more information on using your API Token with the Dataverse APIs.

Expiration Date

2022-01-21

f17fd68b-3b36-4b2e-8173-8d6fd14f98d4

Recreate Token

Revoke Token

User Guide

Admin Guide

API Guide

Installation Guide

Developer Guide

Style Guide



guides.dataverse.org

Step 3a: Create a Dataverse Collection: Json File

Download **dataverse-complete.json** file and modify it to suit your needs.

```
{
  "name": "Scientific Research",
  "alias": "science",
  "dataverseContacts": [
    {
      "contactEmail": "pi@example.edu"
    },
    {
      "contactEmail": "student@example.edu"
    }
  ],
  "affiliation": "Scientific Research University",
  "description": "We do all the science.",
  "dataverseType": "LABORATORY"
}
```

Step 3b: Create a Dataverse Collection: Execute a Curl Command

```
curl -H X-Dataverse-key:xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxxx -X POST https://demo.dataverse.org/api/dataverses/root --upload-file dataverse-complete.json
```

Step 4a: Create a Dataset: Json File

As a starting point, you can download **dataset-finch1.json** and modify it to meet your needs.

```
1  ✓ {  
2  ✓   "datasetVersion": {  
3  ✓     "metadataBlocks": {  
4  ✓       "citation": {  
5  ✓         "fields": [  
6  ✓           {  
7  ✓             "value": "Darwin's Finches",  
8  ✓             "typeClass": "primitive",  
9  ✓             "multiple": false,  
10 ✓             "typeName": "title"  
11 ✓           },  
12 ✓           {  
13 ✓             "value": [  
14 ✓               {  
15 ✓                 "authorName": {  
16 ✓                   "value": "Finch, Fiona",  
17 ✓                   "typeClass": "primitive",  
18 ✓                   "multiple": false,  
19 ✓                   "typeName": "authorName"  
20 ✓                 },  
21 ✓                 "authorAffiliation": {  
22 ✓                   "value": "Birds Inc.",  
23 ✓                   "typeClass": "primitive",  
24 ✓                   "multiple": false,  
25 ✓                   "typeName": "authorAffiliation"  
26 ✓                 }  
27 ✓               ]  
28 ✓             },  
29 ✓             "typeClass": "compound",  
30 ✓             "multiple": true,  
31 ✓             "typeName": "author"
```

Step 4b: Create a Dataset: Execute Curl Command, Note the PID

```
curl -H "X-Dataverse-key:xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxxx" -X POST "https://demo.dataverse.org/api/dataverses/root/datasets" --upload-file "dataset-finch1.json"
```

You should expect an HTTP 200 (“OK”) response and JSON indicating the database ID and Persistent ID (PID such as DOI or Handle) that has been assigned to your newly created dataset.

Step 5: Add Files: Execute Curl Command, Repeat as Needed

```
curl -H X-Dataverse-key:xxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx -X POST -F file=@data.tsv  
-F 'jsonData={"description":"My description.", "directoryLabel":"data/subdir1", "categories":["Data"], "restrict":"false"}' "https://demo.dataverse.org/api/datasets/:persistentId/add?persistentId=doi:10.5072/FK2/J8SJZB"
```

- A description of the file
- The “File Path” of the file, indicating which folder the file should be uploaded to within the dataset
- Whether or not the file is restricted

```
curl -H X-Dataverse-key:xxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx -X POST -F file=@data.tsv  
-F "https://demo.dataverse.org/api/datasets/:persistentId/add?persistentId=doi:10.5072/FK2/J8SJZB"
```

Step 6a: Publish Dataverse Collection

```
curl -H X-Dataverse-key:xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxxx -X POST https://demo.data  
verse.org/api/datasets/root/actions/:publish
```


Step 6b: Publish Dataset

```
curl -H "X-Dataverse-key: xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx" -X POST "https://demo.dataverse.org/api/datasets/:persistentId/actions/:publish?persistentId=doi:10.5072/FK2/J8SJZB&type=major"
```

Your Dataset has been Published!

Recap!

- Create an Account (in the UI) ✓
- Create an API Token (in the UI) ✓
- Create a Dataverse collection (optional, but recommended!) ✓
- Create a Dataset ✓
- Add File(s) ✓
- Publish your Dataset (and Dataverse collection) ✓

A few last things before the Demo

- [Guides.dataverse.org](https://guides.dataverse.org) is your friend!
- We've only scratched the surface of what Dataverse Software APIs can do
- We try to make everything you can do in the UI available through the APIs
- SWORD API is a standard implementation - we focused on the Native API today

Follow this link to the demo website!

<https://demo.dataverse.org/dataverse/datafest2021>

Thank You!

The Dataverse Project: dataverse.org

Harvard Dataverse: dataverse.harvard.edu

Test the features for Dataverse at: demo.dataverse.org

Contact support@dataverse.harvard.edu with questions, or ask the Dataverse Project Google Group:

<https://groups.google.com/g/dataverse-community>