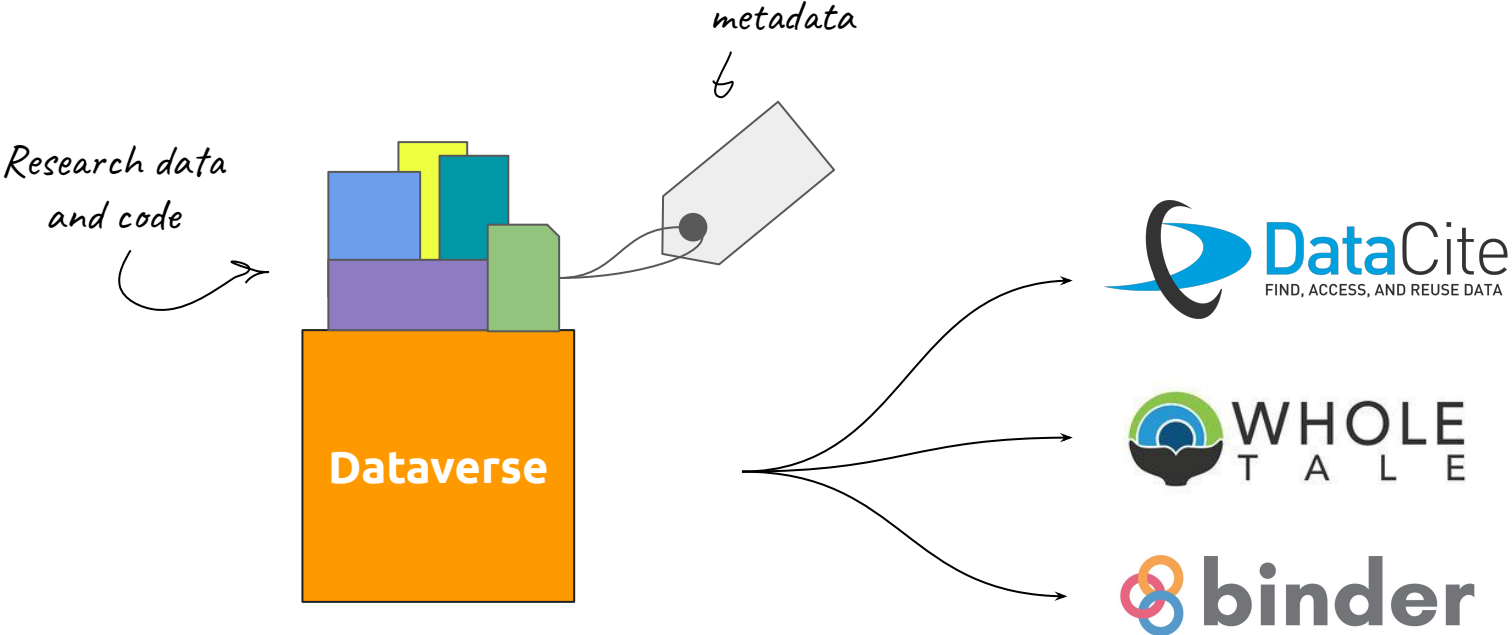


# Dataverse Integration with GitHub via a GitHub Action

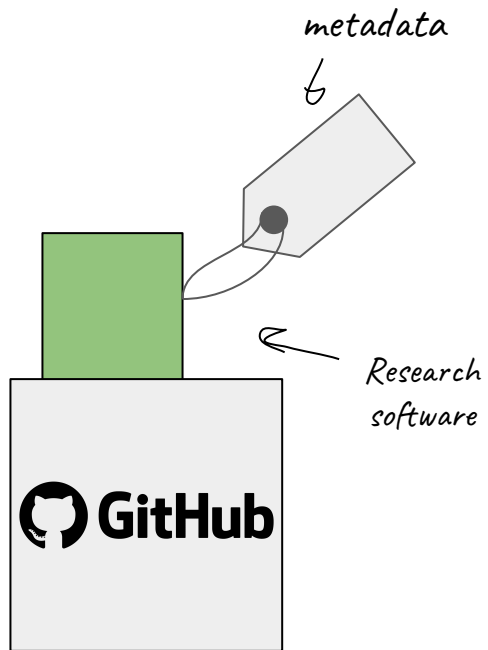
Dataverse Community Meeting 2022  
June 14, 2022

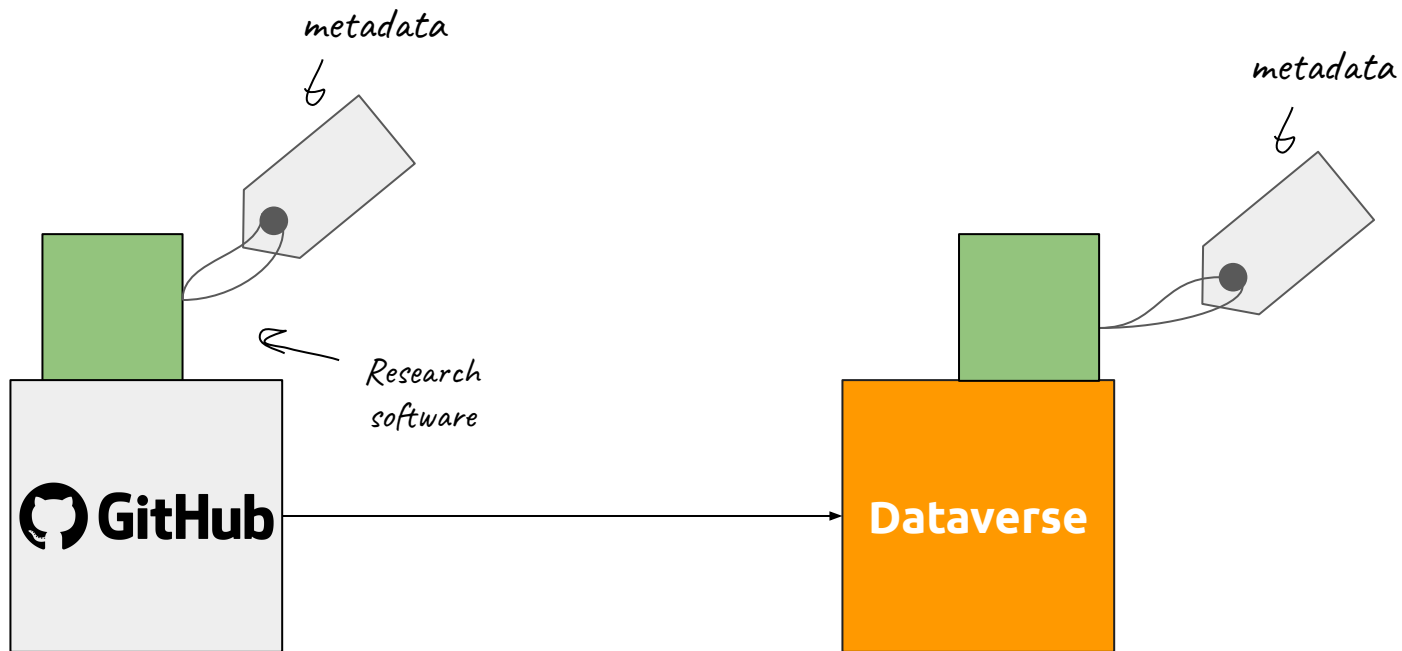
Ana Trisovic, Research Associate  
Harvard Biostatistics & the Institute for Quantitative Social Science

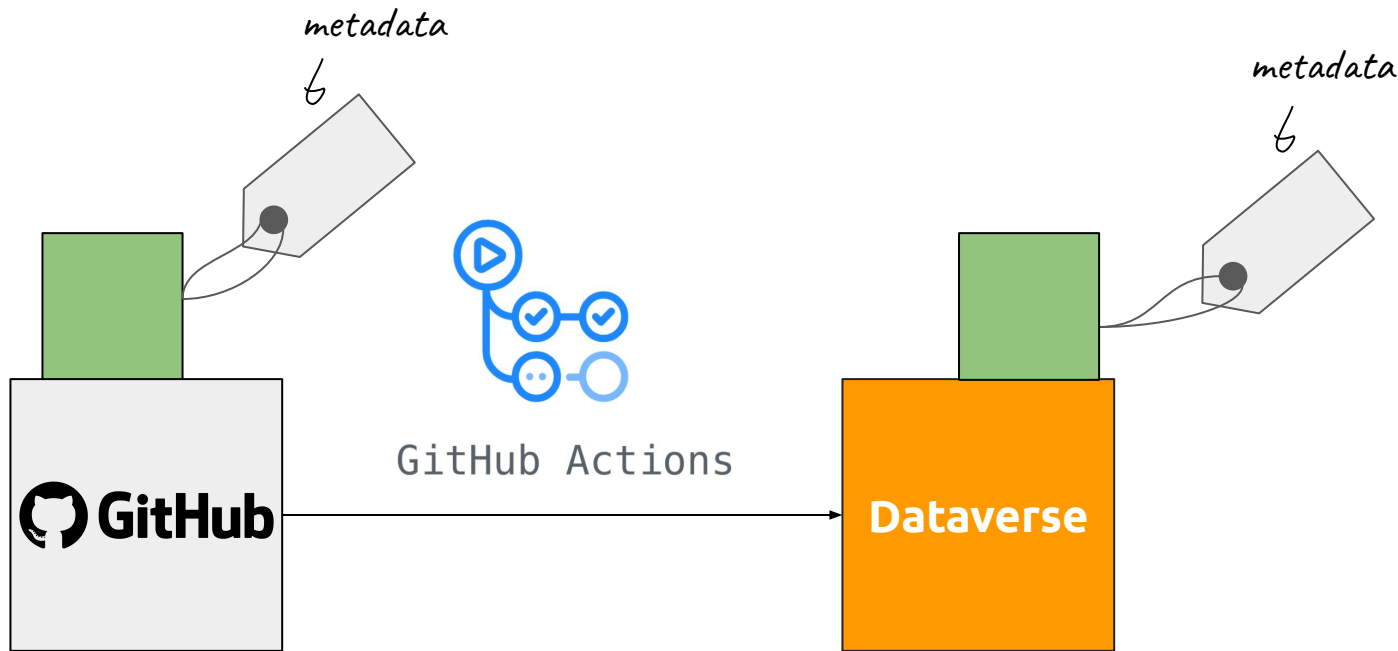
# Dataverse integrations

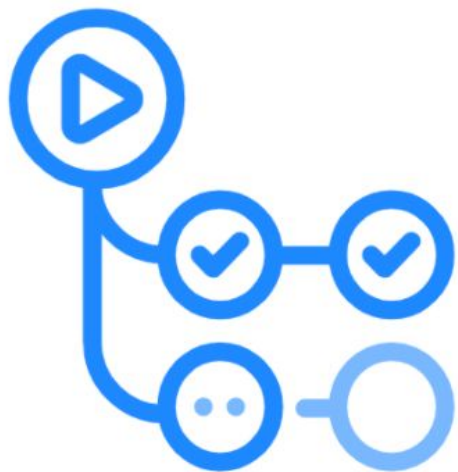


...









# GitHub Actions

- a continuous integration and continuous delivery (CI/CD) platform
- automates build, test, and deployment pipeline
- a composite workflow that can be used and reused in multiple git repositories

---



GitHub Action

# Dataverse Uploader Action

v1.1 Latest version

Use latest version

## Dataverse Uploader

This action automatically uploads GitHub repository content to a Dataverse dataset. It can upload the entire repository or its subdirectory into an existing dataset on a target Dataverse installation. The action is customizable, allowing you to fully replace a dataset, add to the dataset, publish it or leave it as a draft version on Dataverse.

The action provides some additional metadata to the dataset, such as the origin GitHub repository, and it preserves the directory tree structure.

### Input parameters

To use this action, you will need the following input parameters:

Parameter	Required	Description
		This is your personal access token that you can

Stars

☆ Star 7

Contributors



Categories

Publishing Utilities

Links

📄 IQSS/dataverse-uploader

🗨️ Open issues 0

🔗 Pull requests 0

🗨️ Report abuse

# How to use the Dataverse Uploader GitHub Action?



Demo Dataverse >

# Dataverse GitHub Action Test

Version 2.0



Trisovic, Ana, 2022, "Dataverse GitHub Action Test", <https://doi.org/10.70122/FK2/ALROOH>, Demo Dataverse, V2, UNF:6:wedOIQYVA+QCjs3zlEhegg== [fileUNF]

[Cite Dataset](#) ▾

[Learn about Data Citation Standards.](#)

[Access Dataset](#) ▾

[Edit Dataset](#) ▾

[Link Dataset](#)

[Contact Owner](#)

[Share](#)


Dataset Metrics 

0 Downloads 



<> Code Issues Pull requests Actions

master 1 branch 0 tags Go

 atrisovic first commit

README.md	first commit
comments.csv	first commit
issues.csv	first commit
joss_review.ipynb	first commit
requirements.txt	first commit



Dataverse

Demo Dataverse >

## Dataverse GitHub Action Test

Version 2.0

 Trisovic, Ana, 2022, "Dataverse GitHub Action Test", <https://doi.org/10.70122/FK2/ALROOH>, Demo Dataverse, V2, UNF:6:wedOIQYVA+QCjs3zIEhegg== [fileUNF]

[Cite Dataset](#)

[Learn about Data Citation Standards.](#)

[Access Dataset](#)

[Edit Dataset](#)

[Link Dataset](#)

[Contact Owner](#) [Share](#)

Dataset Metrics [?](#)

0 Downloads [?](#)



<> Code Issues Pull requests Action

master 1 branch 0 tags

atrisovic Update action.yml ✓

📁 .github/workflows	Update action.yml
📄 README.md	Update README.md
📄 comments.csv	first commit
📄 issues.csv	first commit
📄 joss_review.ipynb	first commit
📄 requirements.txt	first commit




Dataverse

Demo Dataverse >

## Dataverse GitHub Action Test

Version 2.0

 Trisovic, Ana, 2022, "Dataverse GitHub Action Test", <https://doi.org/10.70122/FK2/ALROOH>, Demo Dataverse, V2, UNF:6:wedOIQYVA+QCjs3zIEhegg== [fileUNF]

[Cite Dataset](#)

[Learn about Data Citation Standards.](#)

[Access Dataset](#)

[Edit Dataset](#)

[Link Dataset](#)

[Contact Owner](#) [Share](#)

Dataset Metrics [?](#)

0 Downloads [?](#)



<> Code Issues Pull requests Action

master 1 branch 0 tags

atrisovic Update action.yml

📁 .github/workflows	Update
📄 README.md	Update
📄 comments.csv	first co
📄 issues.csv	first co
📄 joss_review.ipynb	first co
📄 requirements.txt	first co

Dataverse

Demo Dataverse >

# Dataverse GitHub Action Test

Version 2.0

Trisovic, Ana, 2022, "Dataverse GitHub Action Test", <https://doi.org/10.70122/FK2/ALROOH>, Demo Dataverse, V2, UNF:6:wedOIQYVA+QCjs3zlEhegg== [fileUNF]

Data Citation Standards.

Access Dataset ▾

Edit Dataset ▾

Link Dataset

Share

```
/.github / workflows / action.yml
1 name: Push repo to Dataverse
2
3 on:
4   push:
5     workflow_dispatch:
6
7 jobs:
8   build:
9     runs-on: ubuntu-latest
10    steps:
11      - name: Send repo to Dataverse
12        uses: IQSS/dataverse-uploader@v1.2
13        with:
14          DATAVERSE_TOKEN: ${secrets.DATVERSE_TOKEN}
15          DATAVERSE_SERVER: https://demo.dataverse.org
16          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROOH
17          DELETE: True
18          PUBLISH: True
```



<> Code Issues Pull requests Action

master 1 branch 0 tags

atrisovic Update action.yml

- .github/workflows Update
- README.md Update
- comments.csv first co
- issues.csv first co
- joss\_review.ipynb first co
- requirements.txt first co



```
/.github / workflows / action.yml
1 name: Push repo to Dataverse
2
3 on:
4   push:
5     workflow_dispatch:
6
7 jobs:
8   build:
9     runs-on: ubuntu-latest
10    steps:
11      - name: Send repo to Dataverse
12        uses: IQSS/dataverse-uploader@v1.2
13        with:
14          DATAVERSE_TOKEN: ${secrets.DATVERSE_TOKEN}
15          DATAVERSE_SERVER: https://demo.dataverse.org
16          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALR00
17          DELETE: True
18          PUBLISH: True
```

Dataverse

Demo Dataverse >

# Dataverse GitHub Action Test

Version 2.0

Trisovic, Ana, 2022, "Dataverse GitHub Action Test", <https://doi.org/10.70122/FK2/ALR00H>, Demo Dataverse, V2, UNF:6:wedOIQYVA+QCjs3zIEhegg== [fileUNF]

Data Citation Standards.

Access Dataset

Edit Dataset

Link Dataset

Share



Dataverse



GitHub events that trigger  
the action.

### `/.github/workflows/action.yml`

```
1 name: Push repo to Dataverse
2
3 on:
4   push:
5   workflow_dispatch:
6
7 jobs:
8   build:
9     runs-on: ubuntu-latest
10    steps:
11      - name: Send repo to Dataverse
12        uses: IQSS/dataverse-uploader@v1.2
13        with:
14          DATAVERSE_TOKEN: ${secrets.DATAVERSE_TOKEN}
15          DATAVERSE_SERVER: https://demo.dataverse.org
16          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROO
17          GITHUB_DIR: docs/slides
18          DELETE: True
19          PUBLISH: True
```

## /.github/workflows/action.yml

```
1 name: Push repo to Dataverse
2
3 on:
4   push:
5   workflow_dispatch:
6
7 jobs:
8   build:
9     runs-on: ubuntu-latest
10    steps:
11      - name: Send repo to Dataverse
12        uses: IQSS/dataverse-uploader@v1.2
13        with:
14          DATAVERSE_TOKEN: ${secrets.DATVERSE_TOKEN}
15          DATAVERSE_SERVER: https://demo.dataverse.org
16          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROO
17          GITHUB_DIR: docs/slides
18          DELETE: True
19          PUBLISH: True
```

GitHub events that trigger the action.

Required arguments (Dataverse token, server, dataset DOI)

## `/.github/workflows/action.yml`

```
1 name: Push repo to Dataverse
2
3 on:
4   push:
5   workflow_dispatch:
6
7 jobs:
8   build:
9     runs-on: ubuntu-latest
10    steps:
11      - name: Send repo to Dataverse
12        uses: IQSS/dataverse-uploader@v1.2
13        with:
14          DATAVERSE_TOKEN: ${secrets.DATVERSE_TOKEN}
15          DATAVERSE_SERVER: https://demo.dataverse.org
16          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROO
17          GITHUB_DIR: docs/slides
18          DELETE: True
19          PUBLISH: True
```

GitHub events that trigger the action.

Required arguments (Dataverse token, server, dataset DOI)

Upload a subdirectory or the whole repository



## `/.github/workflows/action.yml`

```
1 name: Push repo to Dataverse
2
3 on:
4   push:
5   workflow_dispatch:
6
7 jobs:
8   build:
9     runs-on: ubuntu-latest
10    steps:
11      - name: Send repo to Dataverse
12        uses: IQSS/dataverse-uploader@v1.2
13        with:
14          DATAVERSE_TOKEN: ${secrets.DATVERSE_TOKEN}
15          DATAVERSE_SERVER: https://demo.dataverse.org
16          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROO
17          GITHUB_DIR: docs/slides
18          DELETE: True
19          PUBLISH: True
```

GitHub events that trigger the action.

Required arguments (Dataverse token, server, dataset DOI)

Upload a subdirectory **or** the whole repository

Dataset is appended **or** synced up (replaced) upon upload

## `/.github/workflows/action.yml`

```
1 name: Push repo to Dataverse
2
3 on:
4   push:
5   workflow_dispatch:
6
7 jobs:
8   build:
9     runs-on: ubuntu-latest
10    steps:
11      - name: Send repo to Dataverse
12        uses: IQSS/dataverse-uploader@v1.2
13        with:
14          DATAVERSE_TOKEN: ${secrets.DATVERSE_TOKEN}
15          DATAVERSE_SERVER: https://demo.dataverse.org
16          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROO
17          GITHUB_DIR: docs/slides
18          DELETE: True
19          PUBLISH: True
```

GitHub events that trigger the action.

Required arguments (Dataverse token, server, dataset DOI)

Upload a subdirectory **or** the whole repository

Dataset is appended **or** synced up (replaced) upon upload

Dataset is a draft **or** published as a new version

Mandatory

```
DATAVERSE_TOKEN: ${secrets.DATAVERSE_TOKEN}  
DATAVERSE_SERVER: https://demo.dataverse.org  
DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROO
```

Optional

```
GITHUB_DIR: docs/slides  
DELETE: True  
PUBLISH: True
```

**Implementation**

# Implementation

The action uses **pyDataverse** to communicate with the target Dataverse installation

By default:

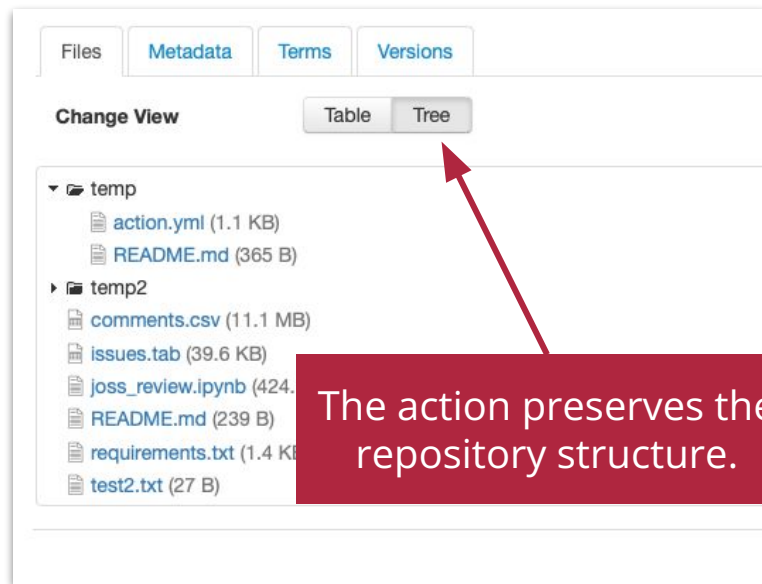
- It uploads the whole git repository to the Dataverse dataset
- It replaces (deletes files) in the Dataverse dataset (to avoid duplication) and uploads the updated content
- It does not publish the updated dataset, but keeps it as a *draft* in Dataverse

# Implementation

The action uses **pyDataverse** to communicate with the target Dataverse installation

By default:

- It uploads the whole git repository to the Dataverse dataset
- It replaces (deletes files) in the Dataverse dataset (to avoid duplication) and uploads the updated content
- It does not publish the updated dataset, but keeps it as a *draft* in Dataverse

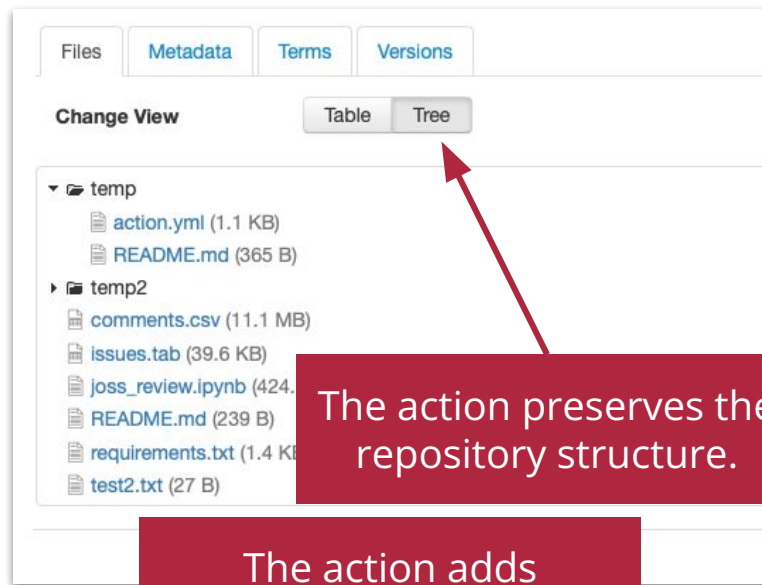


# Implementation

The action uses **pyDataverse** to communicate with the target Dataverse installation

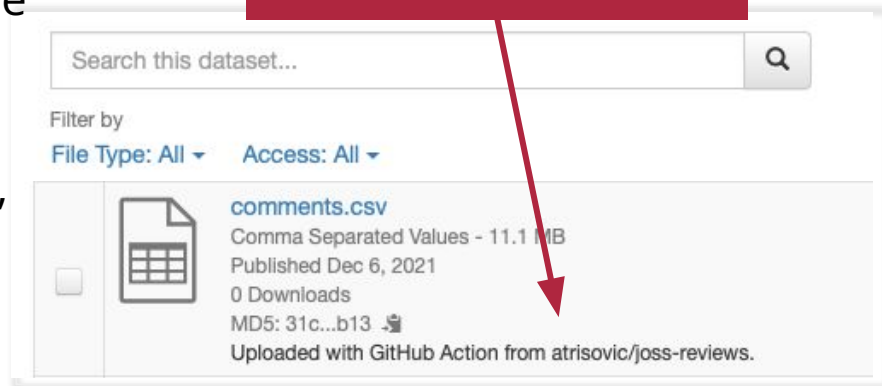
By default:

- It uploads the whole git repository to the Dataverse dataset
- It replaces (deletes files) in the Dataverse dataset (to avoid duplication) and uploads the updated content
- It does not publish the updated dataset, but keeps it as a *draft* in Dataverse



The action preserves the repository structure.

The action adds metadata to the files.



# References and related projects





guides.dataverse.org/en/latest/admin/integrations.html



About ▾

Community

Best Practices ▾

Software ▾

Contact

User Guide

Admin Guide

Dashboard

External Tools

Managing Harvesting Clients

Managing Harvesting Server and Sets

Metadata Customization

Metadata Export

Dataverse Installation Application Timers

Make Data Count

Integrations

User Administration

Managing Datasets and Dataverse Collections

Solr Search Index

IP Groups

## Getting Data In

A variety of integrations are oriented toward making it easier for your researchers to deposit data into your Dataverse installation.

### GitHub

Dataverse integration with GitHub is implemented via a Dataverse Uploader GitHub Action. It is a reusable, composite workflow for uploading a git repository or subdirectory into a dataset on a target Dataverse installation. The action is customizable, allowing users to choose to replace a dataset, add to the dataset, publish it or leave it as a draft version on Dataverse. The action provides some metadata to the dataset, such as the origin GitHub repository, and it preserves the directory tree structure.

For instructions on using Dataverse Uploader GitHub Action, visit <https://github.com/marketplace/actions/dataverse-uploader-action>

### Dropbox

If your researchers have data on Dropbox, you can make it easier for them to get it into your Dataverse installation by setting the `dataverse.dropbox.key` JVM option described in the [Configuration](#) section of the Installation Guide.


### Open Science Framework (OSF)







The Center for Open Science's Open Science Framework (OSF) is an open source software project that facilitates open collaboration in science research across the lifespan of a scientific project.

For instructions on depositing data from OSF to your Dataverse installation, your researchers can visit <https://help.osf.io/hc/en-us/articles/360019737314-Connect-Dataverse-to-a-Project>

**Use this GitHub Action with your project**  
Add this Action to an existing workflow or create a new one. [View on Marketplace](#)

master 1 branch 3 tags [Go to file](#) [Add file](#) [Code](#)

 **atrisovic** add docs and set publish to false ... 22d180f 3 hours ago 🕒 31 commits

 CONTRIBUTING.md	add docs and set publish to false	3 hours ago
 LICENSE	Create LICENSE	7 months ago
 README.md	add docs and set publish to false	3 hours ago
 action.yml	Update action.yml	6 months ago
 dataverse.py	add docs and set publish to false	3 hours ago
 requirements.txt	first commit	7 months ago

# Dataverse Badge Maker

Dataset DOI

10.70122/FK2/ALROOH

Badge Text

Dataverse DOI

DOI

Dataverse

Badge Color

Orange

Green

Yellow

Red

Blue

Grey

Make Badge

Badge

DOI 10.70122/FK2/ALROOH

HTML

```
<a href="https://www.doi.org/10.70122/FK2/ALROOH"></a>
```

Copy

Markdown

```
[]  
(https://www.doi.org/10.70122/FK2/ALROOH)
```

Copy

# Dataverse Badge Maker

Dataset DOI

10.70122/FK2/ALROOH

Badge Text

Dataverse DOI DOI Dataverse

Badge Color

Orange Green Yellow Red Blue Grey

Make Badge

Badge

DOI 10.70122/FK2/ALROOH

HTML

```
<a href="https://www.doi.org/10.70122/FK2/ALROOH"></a>
```

Copy

Markdown

```
[]  
(https://www.doi.org/10.70122/FK2/ALROOH)
```

Copy

README.md

## Related projects

Check out the following related projects:

### Dataverse Badge

Visit [this page](#) to create a Dataverse DOI Badge for your GitHub repository.

Dataverse DOI 10.70122/FK2/LVJADQ





GitHub Action

# Dataverse Uploader Action

v1.1 Latest version

Code Issues Pull requests Action

master 1 branch 0 tags

atrisovic Update action.yml

.github/workflows	Update action.yml
README.md	Update README.m
comments.csv	first commit
issues.csv	first commit
joss_review.ipynb	first commit
requirements.txt	first commit

```
/.github/workflows/action.yml
1 name: Push repo to Dataverse
2
3 on:
4   push:
5     workflow_dispatch:
6
7 jobs:
8   build:
9     runs-on: ubuntu-latest
10    steps:
11      - name: Send repo to Dataverse
12        uses: IQSS/dataverse-uploader@v1.2
13        with:
14          DATAVERSE_TOKEN: ${secrets.DATVERSE_TOKEN}
15          DATAVERSE_SERVER: https://demo.dataverse.org
16          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROO
17          DELETE: True
18          PUBLISH: True
```

Dataverse

Demo Dataverse >

## Dataverse GitHub Action Test

Version 2.0

Trisovic, Ana, 2022, "Dataverse GitHub Action Test", <https://doi.org/10.70122/FK2/ALROOH>, Demo Dataverse, V2, UNF:6:wedOiQYVA+QCjs3zIEhegg== [fileUNF]

Cite Dataset

Learn about Data Citation Standards.

Access Dataset

Edit Dataset

Link Dataset

Contact Owner Share

set Metrics

downloads

