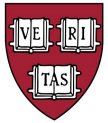
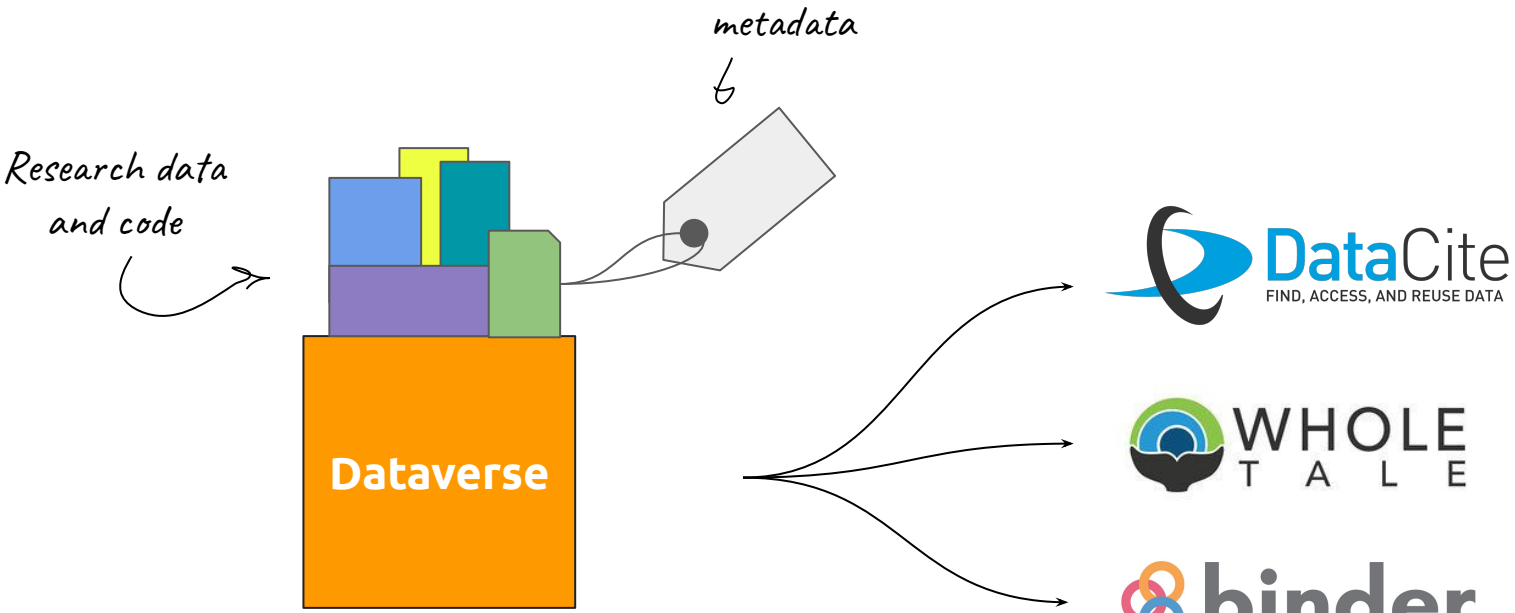


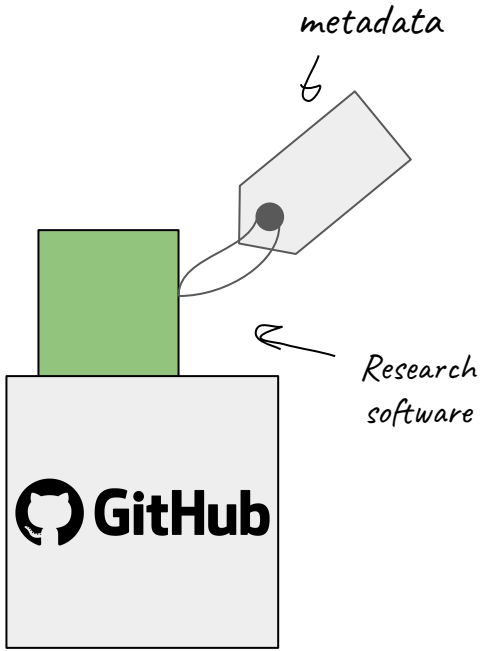
Dataverse Integration with GitHub via a GitHub Action

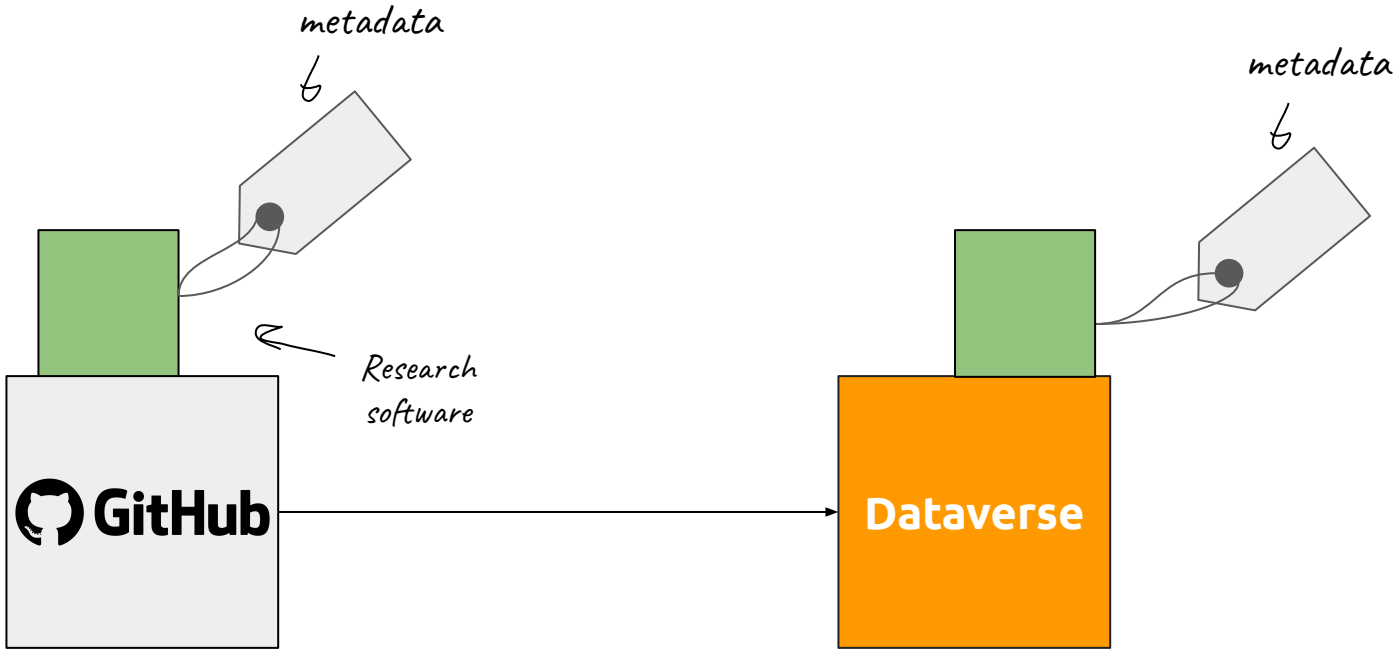
Dataverse Community Meeting 2022
June 14, 2022

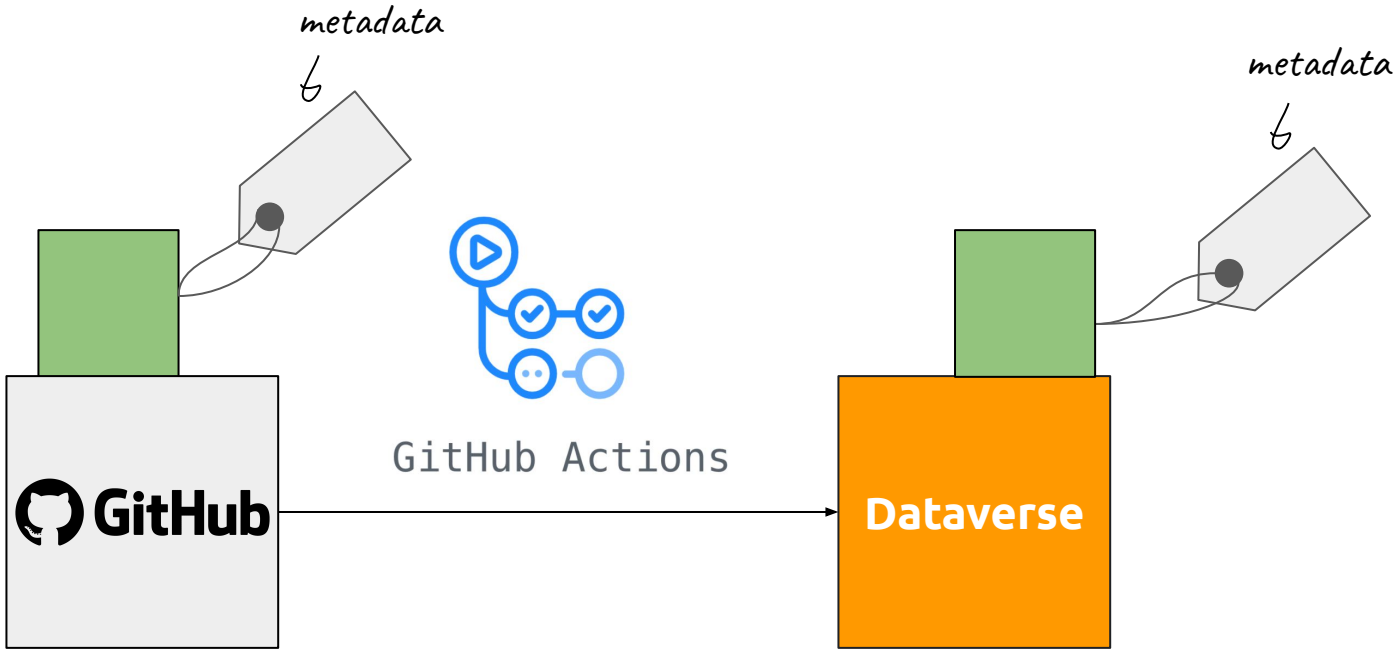
Ana Trisovic, Research Associate
Harvard Biostatistics & the Institute for Quantitative Social Science

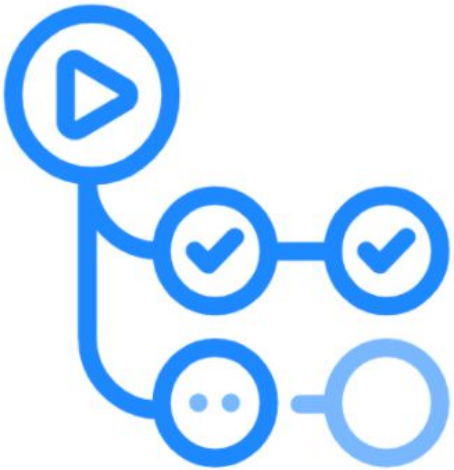
Dataverse integrations




metadata
Research
software
GitHub


metadata
metadata
Research
software
GitHub
Dataverse

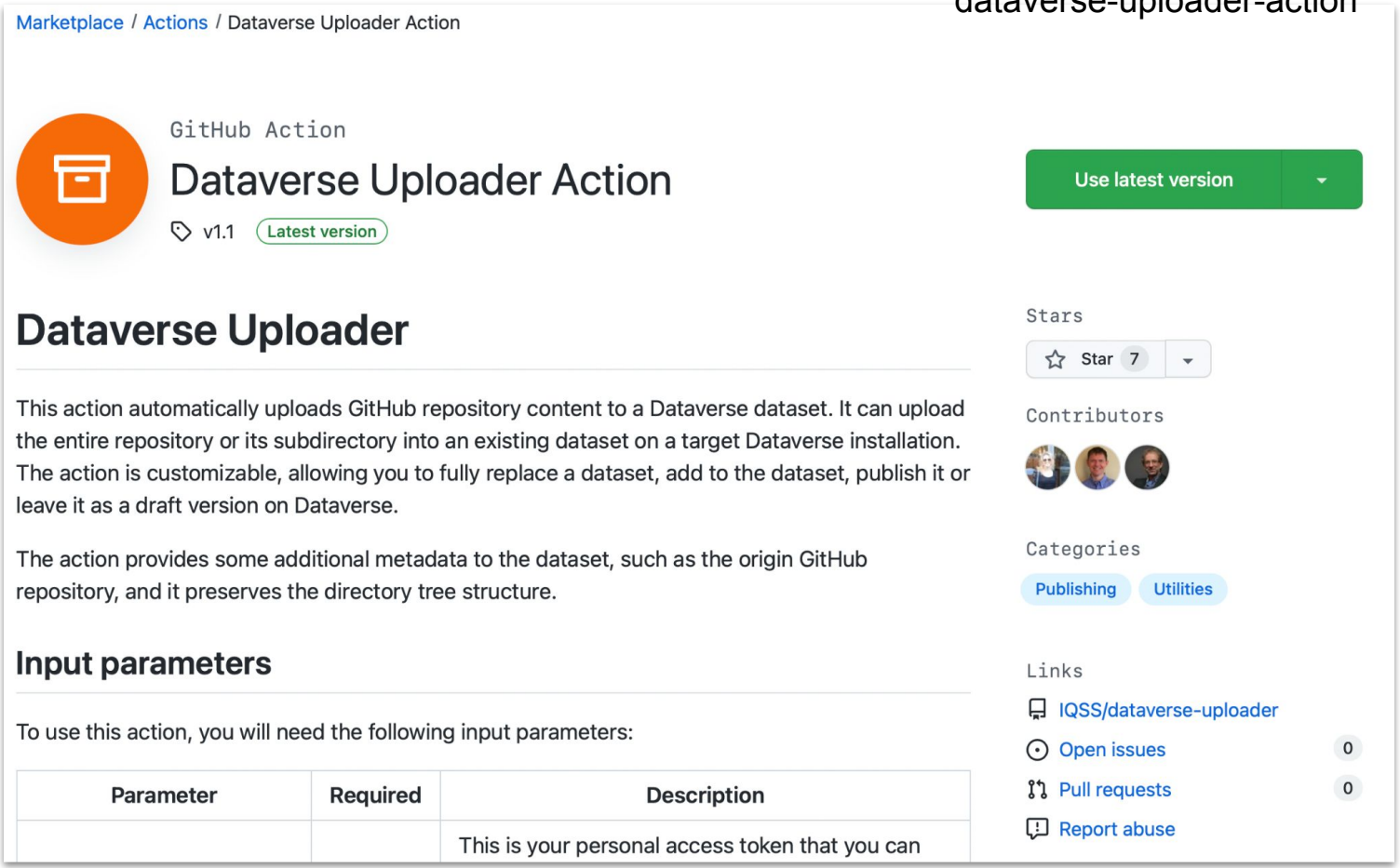

metadata
metadata
GitHub
GitHub Actions
Dataverse



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

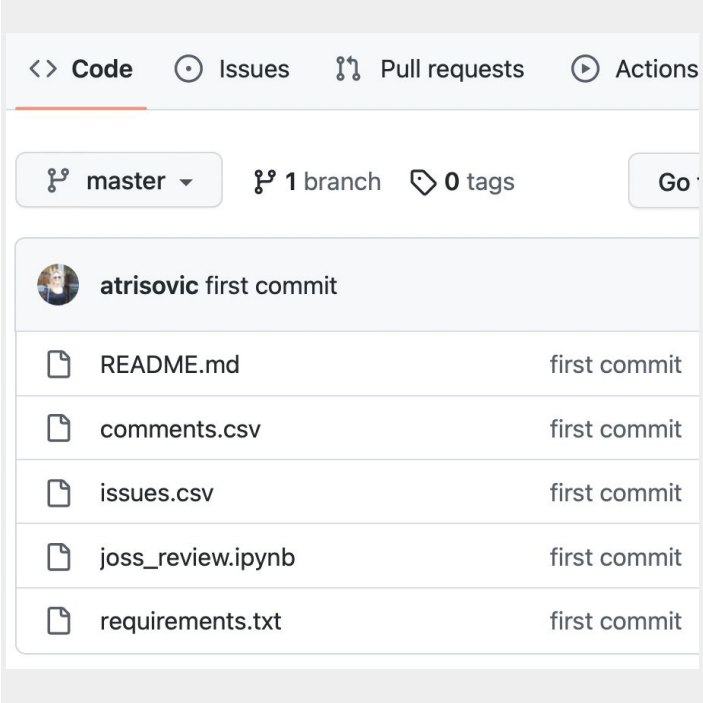
https://github.com/marketplace/actions/dataverse-uploader-action

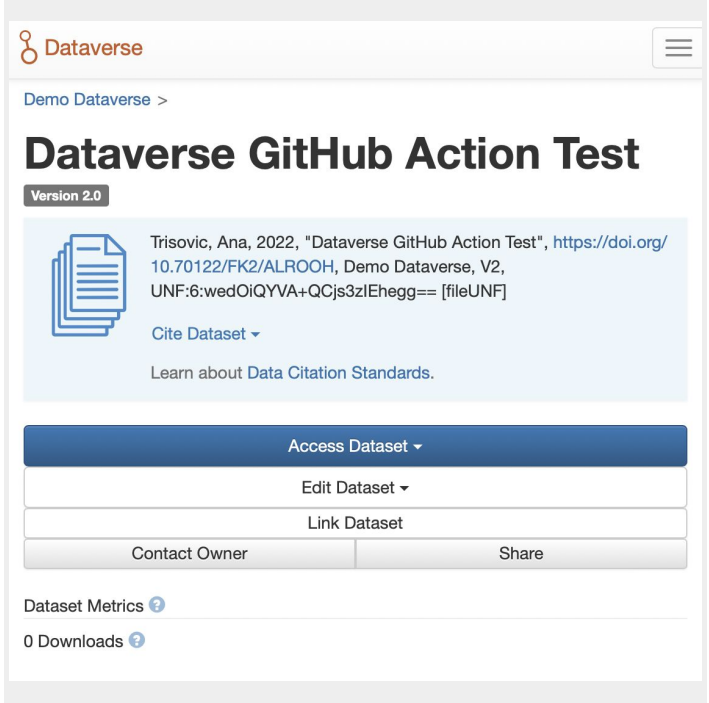


How to use the Dataverse Uploader GitHub Action?

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

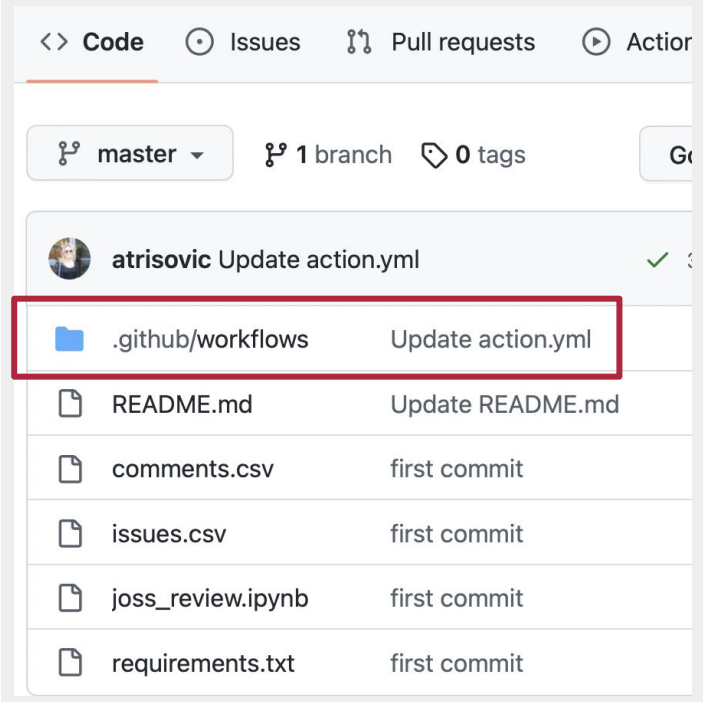
Dataverse

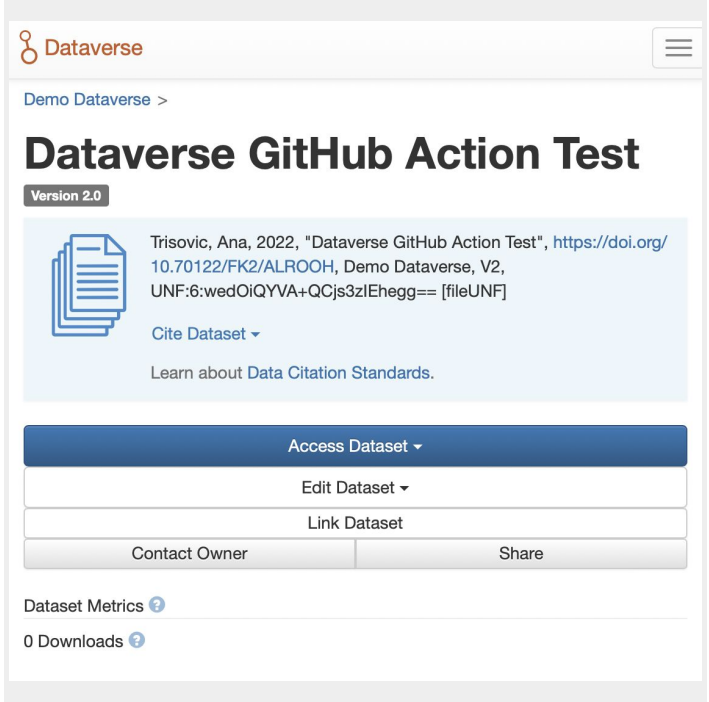




GitHub

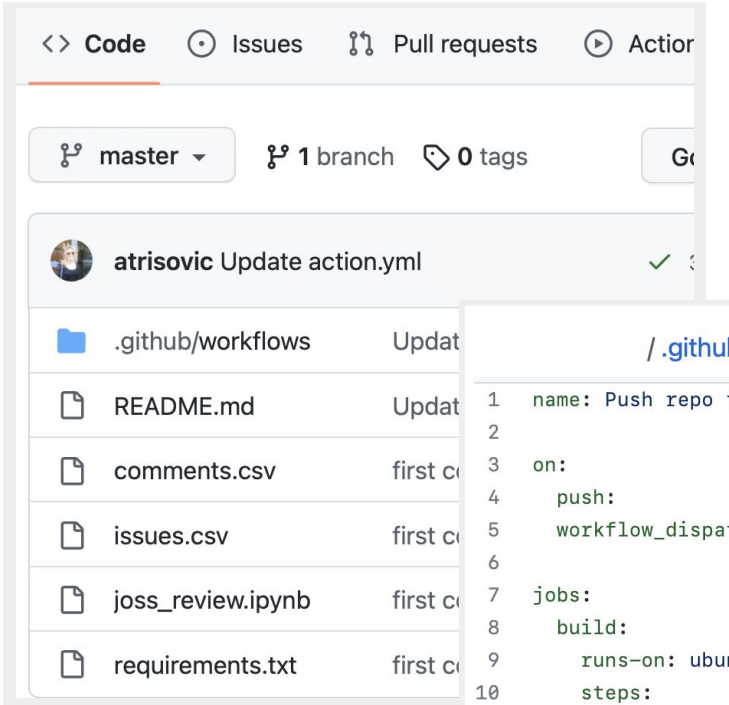
Dataverse

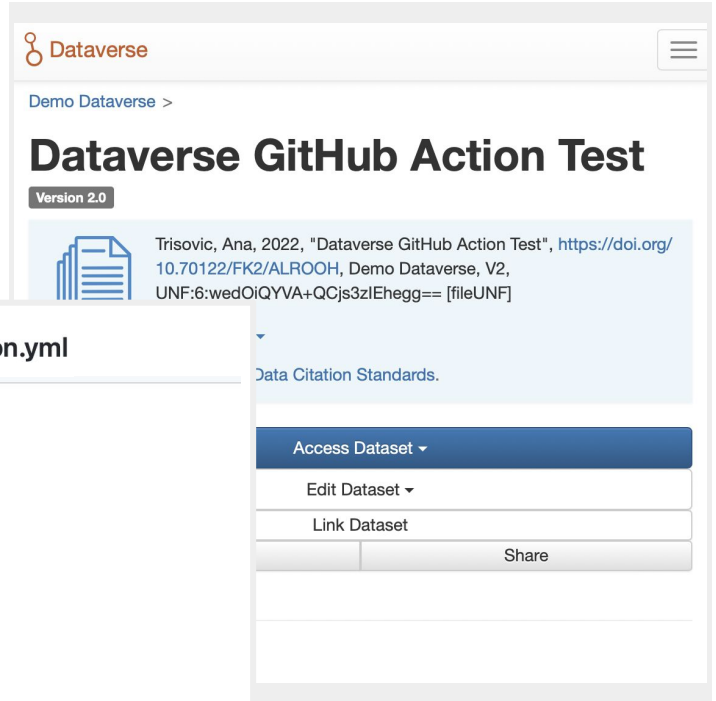




GitHub

Dataverse



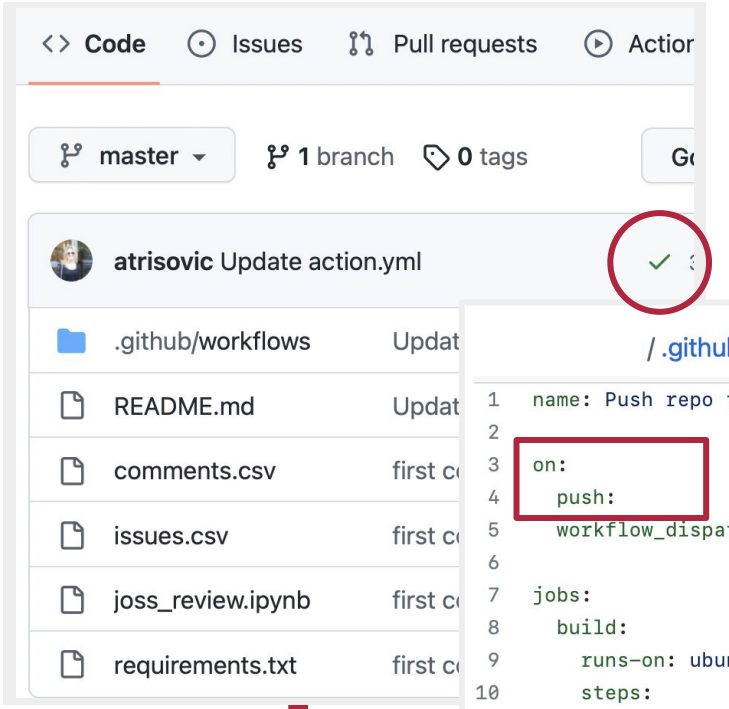


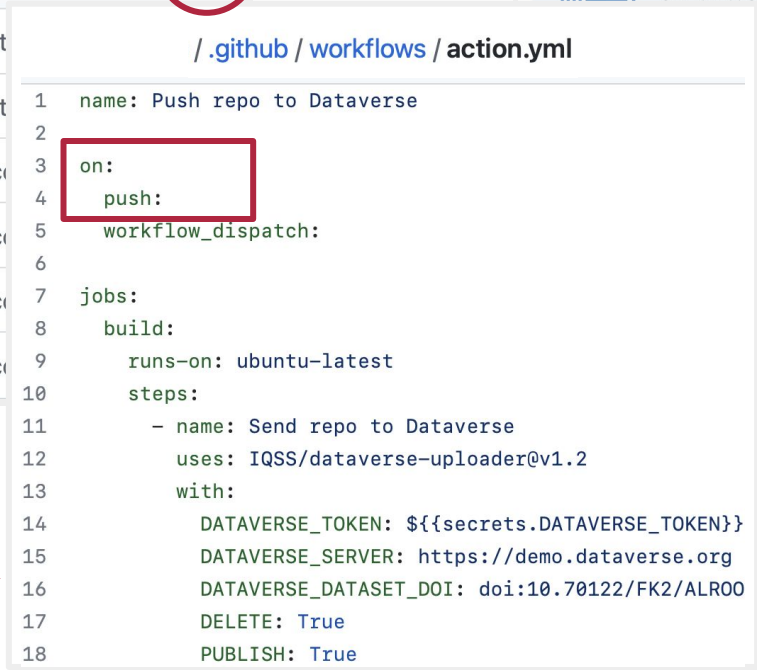
/ .github / workflows / **action.yml**

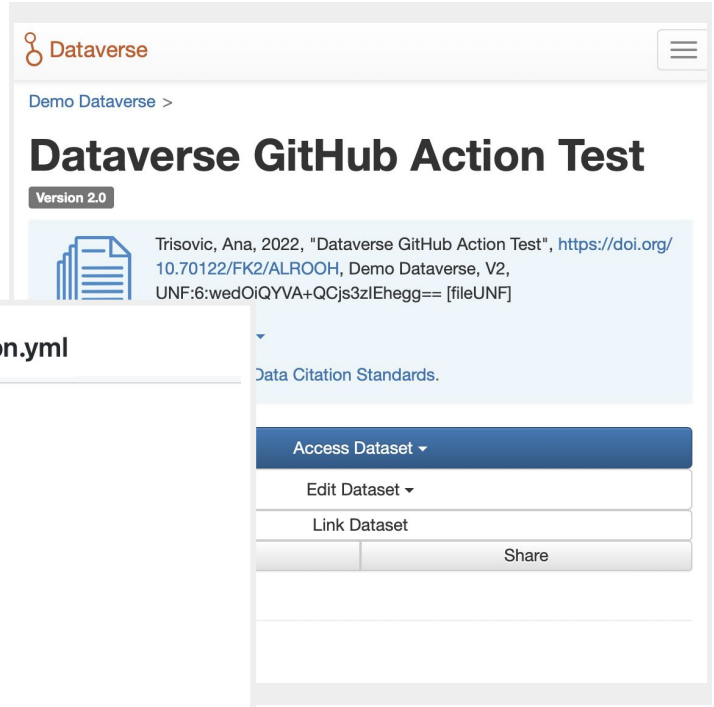
```yaml
name: Push repo to Dataverse

on:
  push:
  workflow_dispatch:

jobs:
  build:
    runs-on: ubuntu-latest
    steps:
      - name: Send repo to Dataverse
        uses: IQSS/dataverse-uploader@v1.2
        with:
          DATAVERSE_TOKEN: ${{secrets.DATAVERSE_TOKEN}}
          DATAVERSE_SERVER: https://demo.dataverse.org
          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROO
          DELETE: True
          PUBLISH: True
```

GitHub

Dataverse







GitHub

Dataverse

GitHub events that trigger the action.

/ .github / workflows / **action.yml**

```yaml
name: Push repo to Dataverse

on:
  push:
  workflow_dispatch:

jobs:
  build:
    runs-on: ubuntu-latest
    steps:
      - name: Send repo to Dataverse
        uses: IQSS/dataverse-uploader@v1.2
        with:
          DATAVERSE_TOKEN: ${{secrets.DATAVERSE_TOKEN}}
          DATAVERSE_SERVER: https://demo.dataverse.org
          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROO
          GITHUB_DIR: docs/slides
          DELETE: True
          PUBLISH: True
```

/ .github / workflows / **action.yml**

```yaml
name: Push repo to Dataverse

on:
  push:
  workflow_dispatch:

jobs:
  build:
    runs-on: ubuntu-latest
    steps:
      - name: Send repo to Dataverse
        uses: IQSS/dataverse-uploader@v1.2
        with:
          DATAVERSE_TOKEN: ${{secrets.DATAVERSE_TOKEN}}
          DATAVERSE_SERVER: https://demo.dataverse.org
          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROO
          GITHUB_DIR: docs/slides
          DELETE: True
          PUBLISH: True
```

GitHub events that trigger the action.

Required arguments (Dataverse token, server, dataset DOI)

/ .github / workflows / **action.yml**

```yaml
name: Push repo to Dataverse

on:
  push:
  workflow_dispatch:

jobs:
  build:
    runs-on: ubuntu-latest
    steps:
      - name: Send repo to Dataverse
        uses: IQSS/dataverse-uploader@v1.2
        with:
          DATAVERSE_TOKEN: ${{secrets.DATAVERSE_TOKEN}}
          DATAVERSE_SERVER: https://demo.dataverse.org
          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROO
          GITHUB_DIR: docs/slides
          DELETE: True
          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**

```yaml
name: Push repo to Dataverse

on:
  push:
  workflow_dispatch:

jobs:
  build:
    runs-on: ubuntu-latest
    steps:
      - name: Send repo to Dataverse
        uses: IQSS/dataverse-uploader@v1.2
        with:
          DATAVERSE_TOKEN: ${{secrets.DATAVERSE_TOKEN}}
          DATAVERSE_SERVER: https://demo.dataverse.org
          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROO
          GITHUB_DIR: docs/slides
          DELETE: True
          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**

```yaml
name: Push repo to Dataverse

on:
  push:
  workflow_dispatch:

jobs:
  build:
    runs-on: ubuntu-latest
    steps:
      - name: Send repo to Dataverse
        uses: IQSS/dataverse-uploader@v1.2
        with:
          DATAVERSE_TOKEN: ${{secrets.DATAVERSE_TOKEN}}
          DATAVERSE_SERVER: https://demo.dataverse.org
          DATAVERSE_DATASET_DOI: doi:10.70122/FK2/ALROO
          GITHUB_DIR: docs/slides
          DELETE: True
          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

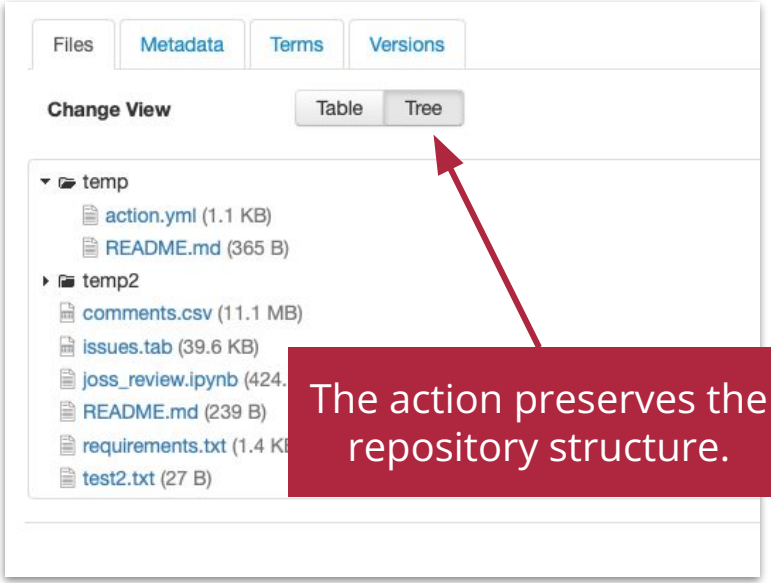
The action uses **pyDataverse** to communicate with the target Dataverse installation
By default:

- 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:

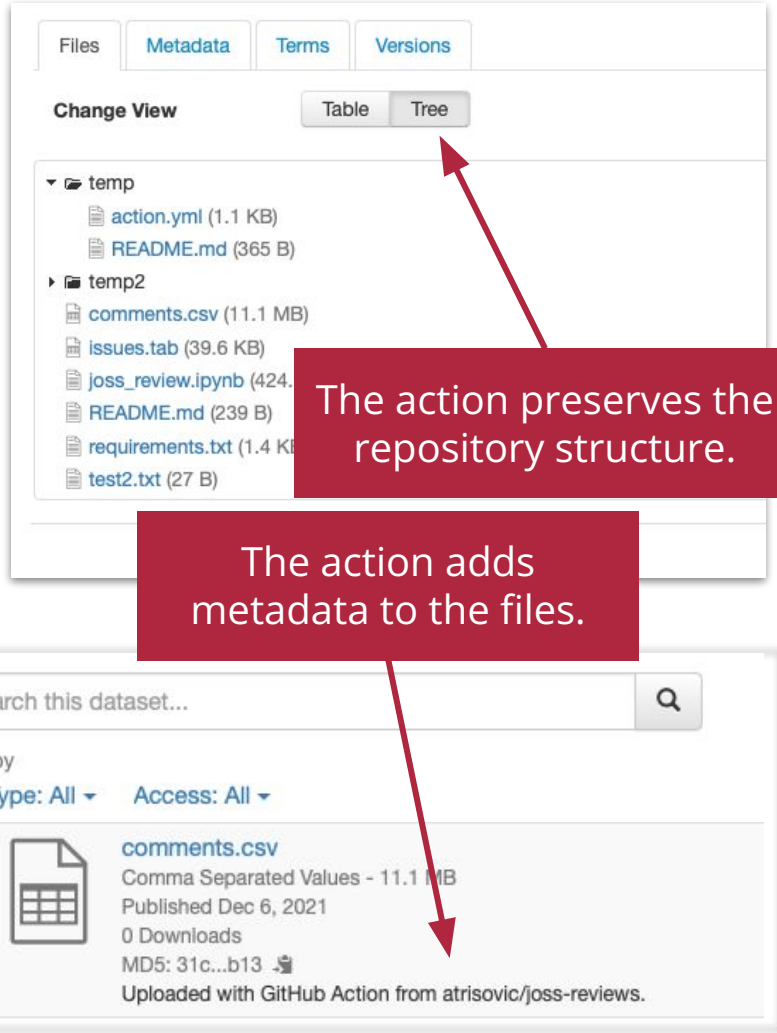
- It uploads the whole git repository to the Dataverse dataset
- It replaces (deletes files) in the Dataverse dataset (to avoid duplication) and uploads the updated content
- It does not publish the updated dataset, but keeps it as a *draft* in Dataverse



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



References and related projects

Search

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

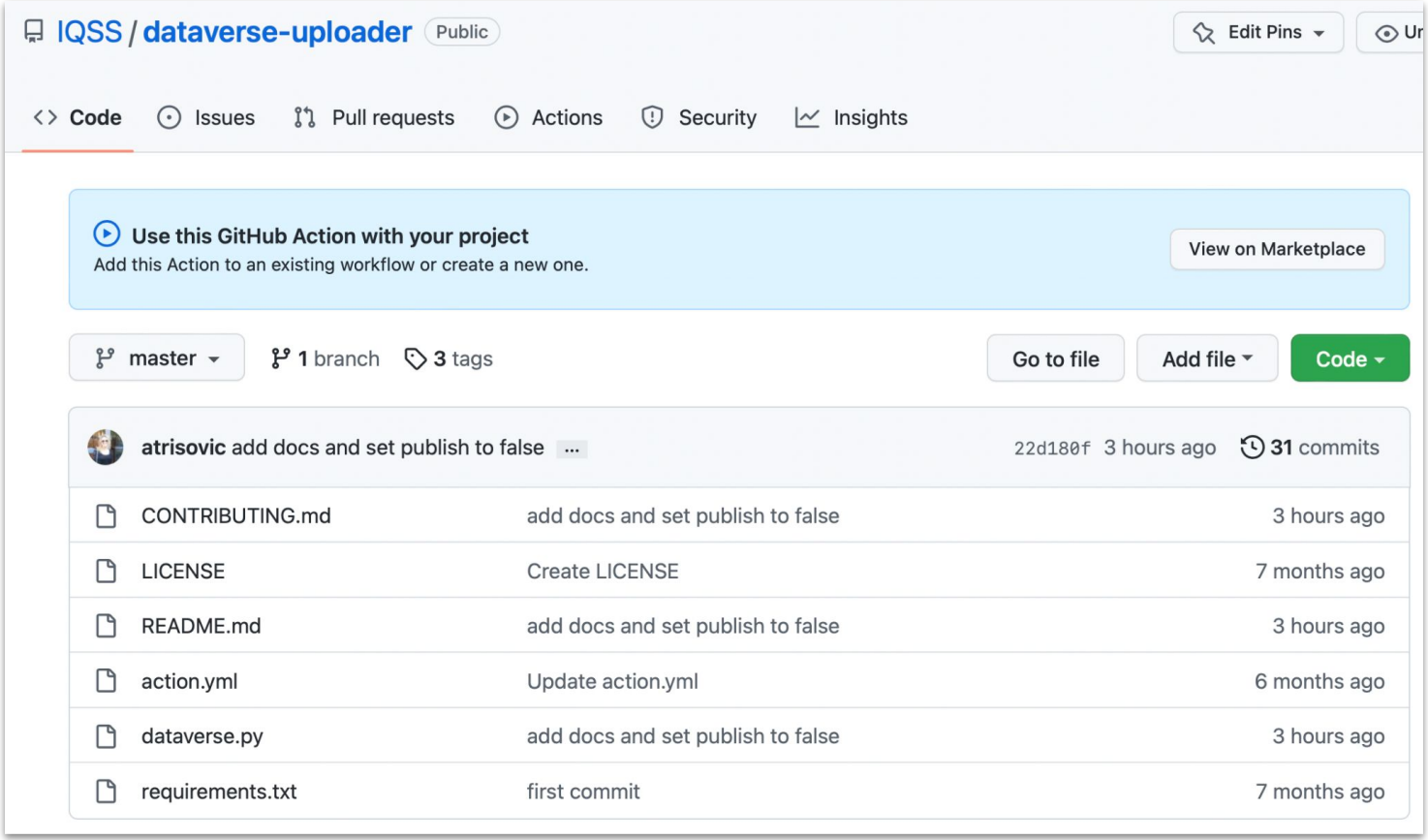
Dropbox

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.

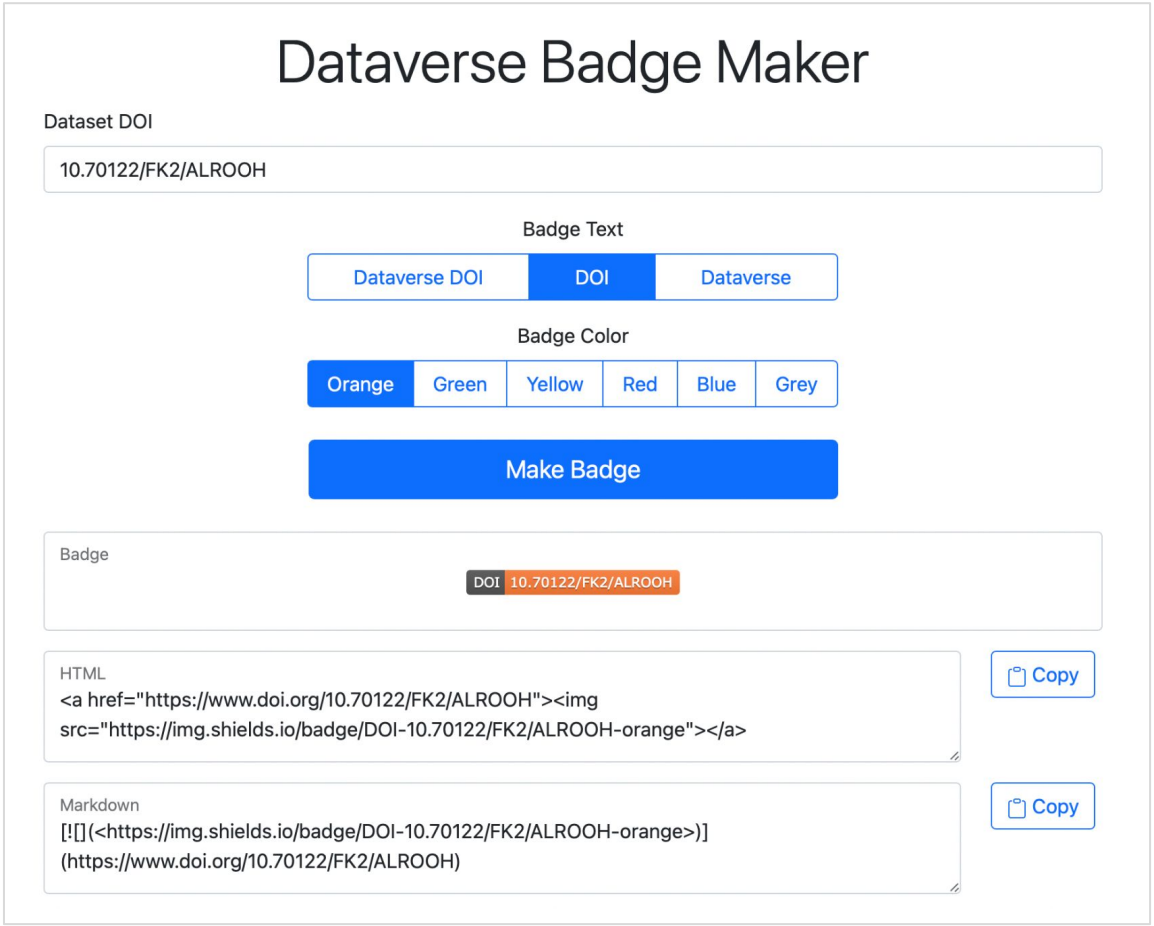
Open Science Framework (OSF)

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



https://github.com/IQSS/dataverse-uploader



https://atrisovic.github.io/dataverse-badge/

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"><img
src="https://img.shields.io/badge/DOI-10.70122/FK2/ALROOH-orange"></a>
```

Copy

Markdown

```
[![](<https://img.shields.io/badge/DOI-10.70122/FK2/ALROOH-orange>)]
(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/LVUADQ

https://atrisovic.github.io/dataverse-badge/

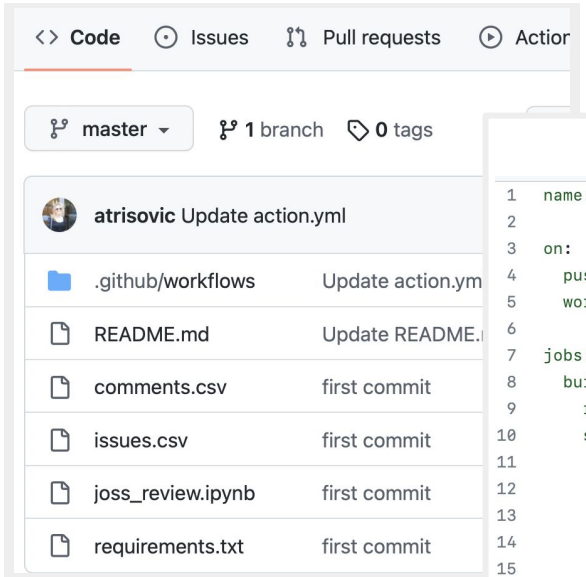
Marketplace / Actions / Dataverse Uploader Action

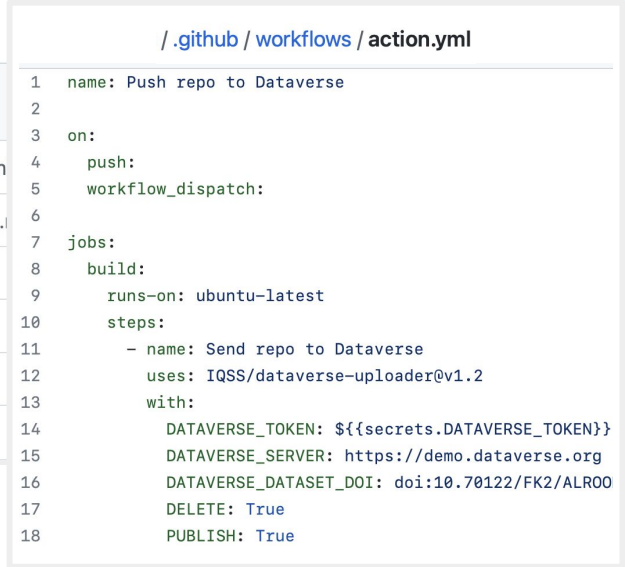
Email: anatrisovic@g.harvard.edu
GitHub & Twitter: atrisovic

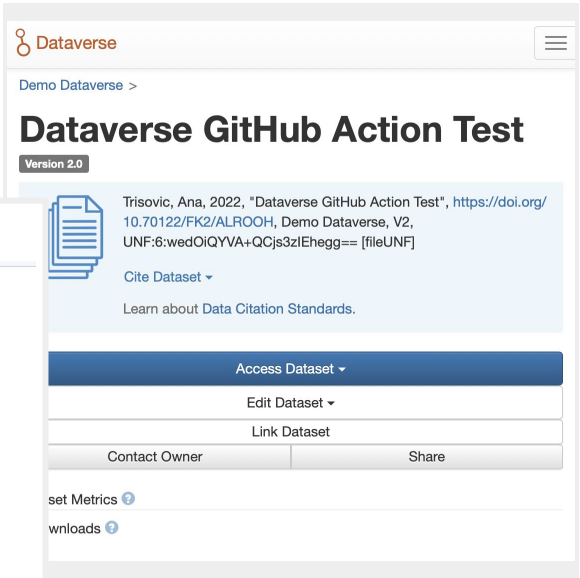
GitHub Action

Dataverse Uploader Action

v1.1 Latest version







GitHub

Dataverse