



Pooch

A friend to fetch your data files

Santiago Soler



www.santisoler.com



@santisoler



Downloading data files

- **Easy** and with **simple code**
- **Reuse** downloaded files
- **Check** if files are **corrupted**



Pooch



fatiando.org/pooch



github.com/fatiando/pooch

- Downloads files
- Caches locally
- Checks integrity



How to use it?

Download a single file

```
import pooch

file_path = pooch.retrieve(
    url="https://awesome.data.com/file.txt",
    known_hash="md5:70e2afd3fd7e336ae478b1e740a5f08e",
)
```

Choosing where to download

```
import pooch

file_path = pooch.retrieve(
    url="https://awesome.data.com/file.txt",
    known_hash="md5:70e2afd3fd7e336ae478b1e740a5f08e",
    path="/home/pycascades/my_data/",
)
```

Different protocols

```
import pooch

file_path = pooch.retrieve(
    url="ftp://basement.org/file.txt",
    known_hash="md5:70e2afd3fd7e336ae478b1e740a5f08e",
)
```

Download files from DOIs

```
import pooch

doi = "10.6084/m9.figshare.13643837"
fname = "australia-ground-gravity.csv"
file_path = pooch.retrieve(
    url=f"doi:{doi}/{fname}",
    known_hash="md5:d47fef200d92c682dc8b63fe31b80364",
)
```

Supports DOIs from **Zenodo**, **figshare** and **Dataverse**

Manage multiple files

```
puppy = pooch.create(  
    path=pooch.os_cache("my-package"),  
    base_url="https://www.somewebpage.org/science/data/",  
    registry={  
        "temperature.csv": "sha256:19uheidh1kjdwhoiwuhc...",  
        "gravity-disturbance.nc": "sha256:1upodh2ioduhw...",  
    },  
)  
  
file_path = puppy.fetch("temperature.csv")
```

Useful for creating **sample datasets** for your package

Modular design

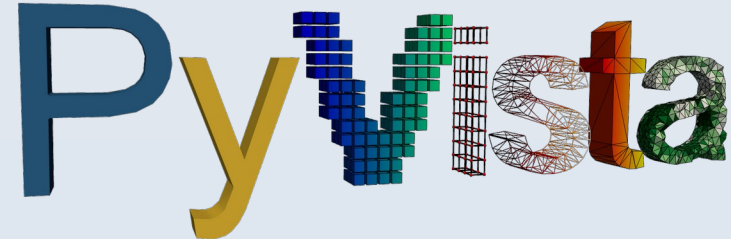
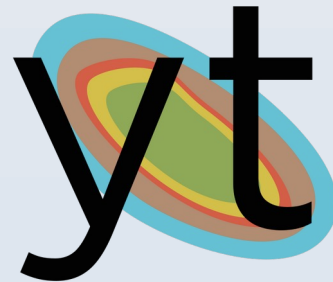
- **Processors** (post-download actions):
 - **Unpack** archives (zip, tar, gz, etc)
- **Downloaders**
 - Add **protocols**
 - **Download** from any **API**



Plug your own classes!



Who uses Pooch?





Check Pooch docs

www.fatiando.org/pooch



www.fatiando.org

Pooch is part of the Fatiando a Terra project

Thank you!



www.fatiando.org/pooch



Slides available under Creative
Commons Attribution International 4.0
License