README

The Dataset\_Repository\_Search.py code is used to search commonly-used general repositories for datasets and software associated with a specific institution.

To modify the code to search for your own institution, the variables in main() need to be edited to the specifics of the institution in question. If certain repositories do not apply (e.g., if your institution lacks an OSF Institutional account), the section can be left blank. The following variables should be modified:

* *ror*: ROR IDs for your institution can be found here: <https://ror.org/>
* osfID: If your institution has an OSF account, the osfID can be found at the end of the base URL for the institution page
* *osf\_ACCESS\_TOKEN*: OSF access tokens can be requested here: <https://osf.io/settings/tokens>
* *zenodoSearchTerms, pangaeaSearchTerms, figshareSearchTerms*, *GBIFSearchTerms:* Specific search terms for each repository can be identified via trial and error in each repository’s search functions
* *zenodo\_ACCESS\_TOKEN*: users should update the Zenodo Access token to one assigned to themselves. Zenodo Access tokens for applications can be requested here: <https://zenodo.org/account/settings/applications/tokens/new/>
* *fileIdentifier*: This variable should be set to the name of your institution and will be used in the file names of all outputs

The file will output individual CSV files for each of the data repositories. Each file will contain a list of DOIs. The lists from each repository can then be subsequently analyzed and combined to remove duplicates.

All dependencies for the code can be found in the requirements.txt file. These can be installed using the following Windows command in the working directory: pip install -r requirements.txt

The swagger client needs to be unzipped in the same directory as the python code.