Catalog Builder

- A “python community package ecosystem” that allows you to generate data catalogs compatible with intake-esm. Available as a Conda package (catalogbuilder)

```
from catalogbuilder.scripts import gen_intake_gfdl
```

- Use it from a Jupyter notebook, a Python script, or from the command-line.
- Catalog Builder GitHub repository
  - Automated build and testing
  - Automated documentation (https://noaa-gfdl.github.io/CatalogBuilder/index.html)

- Cite our work

Radhakrishnan, A., Brown, C., Monge, R., Chang, B., Blanton, C., & Sentman, L. (2024). Catalog Builder for data discovery and analysis at GFDL (Version v03.2024) [Computer software]. https://doi.org/10.5281/zenodo.10787602

Configuration

- Configuration is yaml based. See example here

Example:

- Catalog column names (headers) are set with the **HEADER LIST** variable
- The directory structure and file name expectations are set with the **OUTPUT PATH TEMPLATE** variable

Quickstart

INSTALLATION

1. Install conda package: Ex. conda install noaa-gfdl::catalogbuilder

RUNNING THE CATALOG BUILDER

a. CLI based: gen_intake_gfdl.py --config <path_to_config_file>
   Ex. gen_intake_gfdl.py –config config.yaml
b. From Python: Example python script (and Jupyter notebook friendly)

Catalog builder from Python

```
from catalogbuilder.scripts import gen_intake_gfdl
import sys, os

#This is an example call to run catalog builder using a yaml config
file.
package_dir = os.path.dirname(os.path.abspath(__file__))
configyaml = os.path.join(package_dir, 'configs/config-example.yml')

def create_catalog_from_config(config=configyaml):
    csv, json = gen_intake_gfdl.create_catalog(config=configyaml)
    return(csv,json)
```

Contact

https://github.com/NOAA-GFDL/CatalogBuilder/issues