dmpr Documentation

Release 0.0.2.dev63

Scott Wales

Jan 31, 2018

Contents

Source Code on Github

Contents:

CHAPTER 1

README

1.1 Install

Conda install:

```
conda install -c coecms dmpr
```

Pip install (into a virtual environment):

```
pip install dmpr
```

1.2 Use

Get help:

```
dmpr --help
```

Convert output files to compressed CF-NetCDF:

```
dmpr standardise --model MOM --output cfoutput.nc output1.nc output2.nc
```

1.3 Develop

Development install:

```
git clone https://github.com/coecms/dmpr
cd dmpr
conda env create -f conda/dev-environment.yml
source activate dmpr-dev
pip install -e '.[dev]'
```

```
```

or:

```
module load conda/analysis3
install -e '.[dev]' --user
```

Run tests:

```
py.test
```

Build documentation:

```
python setup.py build_sphinx
```

Upload documentation:

```
git subtree push --prefix docs/_build/html/ origin gh-pages
```

CHAPTER 2

Design

2.1 Requirements

- Use a common tool for all models supported by the Centre
- Support the whole publication pipeline, from model output to publication
- Add metadata at the earliest point possible
- Ensure CF compliance

2.2 Design 1

A CLI tool with multiple commands for different stages in the model's lifetime

- *dmpr post*: Post-process immediately after a model run
- *dmpr stage*: Prepare output for publication, adding metadata from the data management plan and checking metadata compliance
- *dmpr publish*: Final publication steps, adding DOI and moving to final versioned location (admin only?)

2.2.1 Commands

post

The post command has one required argument, the model run directory, and can take an arbitray list of files to post-process

After running the post command the input files have been converted to CF-Compliant NetCDF files and moved to a run-specific output directory

A list of newly created files is printed to standard output

stage

The stage command has one required argument, the job identifier used by post-processing. If a data management plan identifier is not present in the processed files this identifier is also required

The command updates metadata from the DMP and runs a CF compliance check

After running the stage command if the data files pass a CF-Compliance check they are copied to ua8 under their DMP directory

2.2.2 Implementation

cli

The command-line interface is created using Click. Options are kept simple in order to make output consistent between users.

The model is identified either by specifying the name or by reading the files in the run directory, using functions in `dmpr.model`.

models

Each model has its own class, derived from `dmpr.base.Model`. The model must override two functions, *read_configs()* and *post_impl()*, and may optionally override *outfile()* to customise the processed file's name.

read_configs() is passed the run directory, and should read the configuration files held there to set up metadata from the run configuration.

post_impl() is passed the names of the input and output files, and should post-process the input files and write the processed data to the output file.

The base class uses these functions in it's *post()* function, which generates the output path, processes the file and then adds DMP metadata

Linking with a DMP is optional, as it may not be created at the time of the model run. A DMP may be linked after post-processing using *dmpr stage*.

dmp

The *dmpr.dmp.DMP* class holds data management plan related metadata, read from the online database. It has an *addmeta()* function to add metadata it reads from the database to a file, which gets automatically called by the model's *post()* function.

CHAPTER 3

Modules

3.1 dmpr package

3.1.1 Subpackages

dmpr.um package

Submodules

dmpr.um.model module

Module contents

3.1.2 Submodules

3.1.3 dmpr.base module

3.1.4 dmpr.cli module

3.1.5 dmpr.dmp module

class dmpr.dmp.**DMP**(*project*)

Bases: object

DMP related metadata

file_metadata()

Returns metadata that should be added to the file, as a dict

3.1.6 dmpr.model module

3.1.7 dmpr.pkg_info module

3.1.8 Module contents

CHAPTER 4

Indices and tables

- genindex
- modindex
- search

Python Module Index

d

Index

D

F