

# Workforce

Theo Portlock<sup>1</sup>

<sup>1</sup>The University of Auckland

03-04-2024



- Repeatability

# Requirements of reproducibility

- Easy to run/use
  - Good documentation
  - Detailed methods (version control)
  - Multiple platforms/OS
  - Cheap - maximises resources
  - Scalable - as datasets get bigger, this becomes more important
- Easy to understand - (also makes it easier to upgrade)
  - Good tutorials
- Easy to troubleshoot/test
  - Able to resume from previous steps
  - Errors are clear
- Good data hygiene

- Programming Language for creating bioinformatic processing pipelines
- Meets *most* of these requirements
- Containerised, version controlled, and can run on multiple platforms
- Steep learning curve

seqera

Command line Parameters Configuration Datasets Reports

```
nextflow /home/tpor598/biobakery-nextflow/main.nf --project uoa03
```

General

id Zvqp8GjOLqpxjz  
tender\_minsky  
2024-04-29 16:09:00  
-  
d9f96614-b738-44d7-8163-df478036e21f  
tpor598  
/scale\_wlg\_nobackup/filesets/nobackup/uoac  
-  
slurm  
-  
23.10.0 build 5891

Status

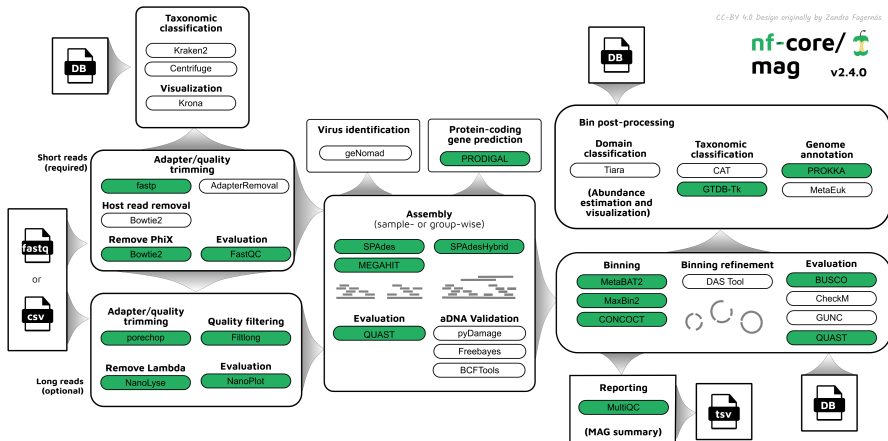
1,096 pending	0 submitted
35 running	0 cached
154 succeeded	15 failed

Processes

KNEADATA\_INIT 1 of 1  
KNFADDDATA 153 of 1,299

Aggregate stats

23 h 11 m 2 s



# Downstream analysis

- 12,000 line R script, python file, or SPSS file
- Collection of random CUSTOM scripts released on github
- Slow
- Often difficult to install correct versions of packages
- Documentation limited to README or methods section, often split between
- Only run on one platform/OS
- ALL THE SAME REQUIREMENTS APPLY!

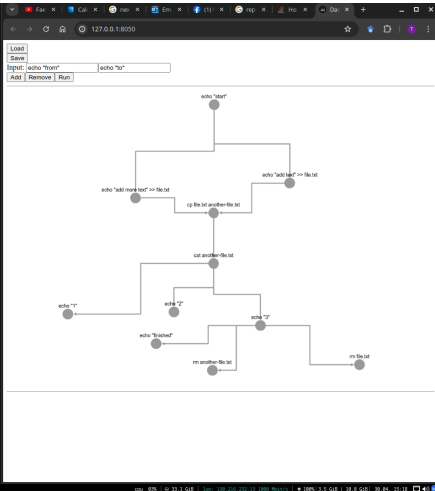


# Workforce

```
[you@hostname] $ theoptheo-20x508dgaq /run/media/theop/maindrive/workforce/workForce(master) wf
Dash is running on http://127.0.0.1:8050/

# Serving Flask app "workforce.gui"
# Debug mode: off
WARNING: This is a development server. Do not use it in a production deployment. Use a producti
# Running on http://127.0.0.1:8050/
Press CTRL+C to quit
Opening in existing browser session
[7392:7392][0108/000000.139355:ERORR:zygote_linux.cc(672)] write: Broken pipe [32]
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET _dash-component-suites/dash/daps/polyfilllib?v=2.11.js
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET _dash-component-suites/dash/daps/react-cv6-v2.11.js
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET _dash-component-suites/dash/daps/react-dom?v=2.11.js
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET _dash-component-suites/dash/cytoscape/dash_cytoscape
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET _dash-component-suites/dash/dash-renderer/build/dash
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET _dash-component-suites/dash/dcc/dash_core_components
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET _dash-component-suites/dash/dcc/dash_core_components
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET _dash-component-suites/dash/dash-table/bundle/v5_2_6
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET _dash-component-suites/dash/html/dash_html_component
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET _dash-layout HTTP/1.1" 200 -
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET _dash-dependencies HTTP/1.1" 200 -
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET _dash-component-suites/dash/dcc/asyncl-upload.js HTTP
127.0.0.1 - - [30/Apr/2024 10:29:55] "GET _Favicon.ico?v=2.11 HTTP/1.1" 200 -
127.0.0.1 - - [30/Apr/2024 15:10:52] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [30/Apr/2024 15:10:52] "GET _dash-component-suites/dash/daps/polyfilllib?v=2.11.js
127.0.0.1 - - [30/Apr/2024 15:10:52] "GET _dash-component-suites/dash/daps/react-cv6-v2.11.js
127.0.0.1 - - [30/Apr/2024 15:10:52] "GET _dash-component-suites/dash/daps/react-dom?v=2.11.js
127.0.0.1 - - [30/Apr/2024 15:10:52] "GET _dash-component-suites/dash/cytoscape/dash_cytoscape
127.0.0.1 - - [30/Apr/2024 15:10:52] "GET _dash-component-suites/dash/dash-renderer/build/dash
127.0.0.1 - - [30/Apr/2024 15:10:52] "GET _dash-component-suites/dash/dcc/dash_core_components
127.0.0.1 - - [30/Apr/2024 15:10:52] "GET _dash-component-suites/dash/dcc/dash_core_components
127.0.0.1 - - [30/Apr/2024 15:10:52] "GET _dash-component-suites/dash/dash-table/bundle/v5_2_6
127.0.0.1 - - [30/Apr/2024 15:10:52] "GET _dash-component-suites/dash/html/dash_html_component
127.0.0.1 - - [30/Apr/2024 15:13:17] "GET _dash-layout HTTP/1.1" 200 -
127.0.0.1 - - [30/Apr/2024 15:13:17] "GET _dash-dependencies HTTP/1.1" 200 -
127.0.0.1 - - [30/Apr/2024 15:13:17] "GET _dash-component-suites/dash/dcc/asyncl-upload.js HTTP
127.0.0.1 - - [30/Apr/2024 15:13:17] "GET _Favicon.ico?v=2.11 HTTP/1.1" 200 -
[you@hostname] $ theoptheo-20x508dgaq /run/media/theop/maindrive/workforce/workForce(master) wf
Dash is running on http://127.0.0.1:8050/

# Serving Flask app "workforce.gui"
# Debug mode: off
WARNING: This is a development server. Do not use it in a production deployment. Use a producti
# WSGI server instead
# Running on http://127.0.0.1:8050/
Press CTRL+C to quit
Opening in existing browser session
127.0.0.1 - - [30/Apr/2024 15:13:38] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [30/Apr/2024 15:13:39] "GET _dash-dependencies HTTP/1.1" 200 -
127.0.0.1 - - [30/Apr/2024 15:13:39] "GET _dash-layout HTTP/1.1" 200 -
127.0.0.1 - - [30/Apr/2024 15:13:39] "GET _dash-component-suites/dash/dcc/asyncl-upload.js HTTP
127.0.0.1 - - [30/Apr/2024 15:13:39] "GET _Favicon.ico?v=2.11 HTTP/1.1" 200 -
```



- A DCG (Directed Cyclic graph) based pipeline designer for downstream analysis
- Installed with 'pip install workforce'
- Integrates into existing tools
- Uses less code
- Process tracking
- Install with pip on any operating system
- Can run over multiple servers in parallel
- Leans on existing command line commands
- Simple interface
- Can interact with a run while it's still running
- Collaborative development

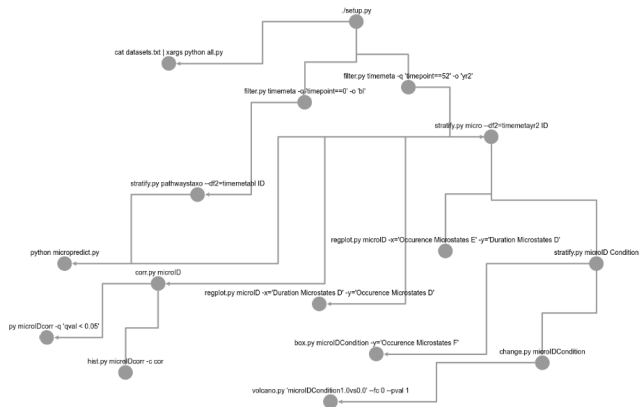
# Workforce - packages

- networkx
- subprocess
- multiprocessing
- dash cytoscape

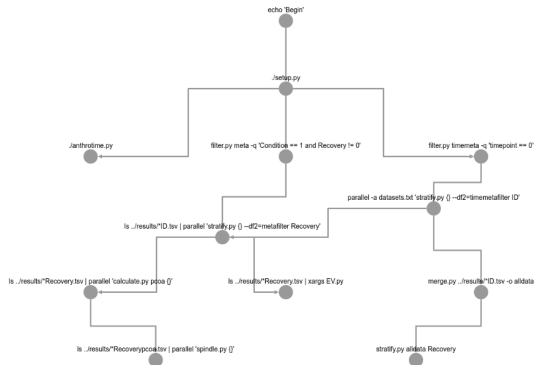
# Workforce - case study

	Well-nourished			MAM			MAM Control
PSS	72	0	66	154	0	145	70
Plasma amino acids	74	0	0	134	136	0	72
Anthropometrics	75	64	0	159	85	0	0
Bayley score	70	0	64	151	0	141	71
fNIRS	14	0	47	58	0	122	0
Glitter wand	70	0	65	152	0	142	72
Plasma lipids	0	0	0	137	137	0	0
EEG microstates	48	0	50	77	0	81	0
Gut pathways	74	74	74	156	147	147	74
EEG PSD	67	0	58	125	0	91	0
Sleep	65	0	65	143	0	139	0
Gut species	74	74	74	156	147	147	74
EEG VEP	36	0	31	78	0	20	0
Plasma vitamins	71	62	0	133	136	136	72
Genetics	75	-	-	150	-	-	74
Wolke's score	70	0	65	152	0	143	71
	0	12	52	0	12	52	104
Weeks after one year baseline							

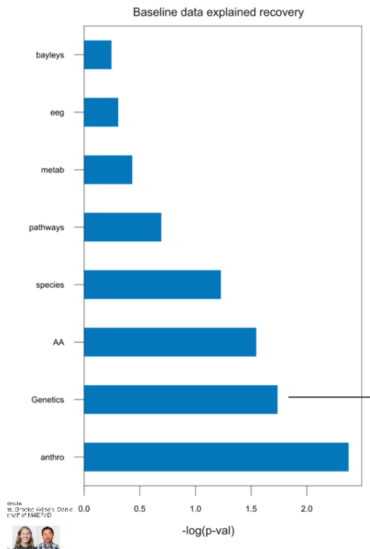
# Workforce - case study



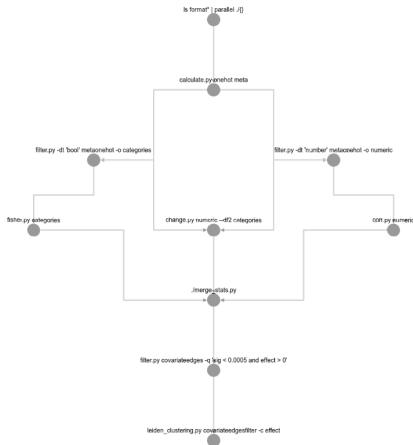
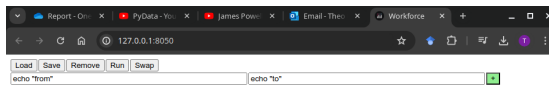
# Workforce - case study



# Workforce - case study



# Workforce - case study



workforce: 2024-05-29 10:58:13.667745



# Workforce - online

The image displays two side-by-side screenshots of the Workforce online IDE, illustrating the workflow from code editing to execution.

**Left Screenshot (Code Editor):**

- The browser address bar shows the URL: `colab.research.google.com/drive/1nSS9MvWqj6t6KHn9pGKKqSLZX...`.
- The file name is `workforce.ipynb`.
- The code editor contains the following Python code:

```
[5] #!workforce
from workforce.gui import gui; gui()
```
- Below the code editor, a GUI diagram is visible. It consists of two gray circular nodes connected by a vertical line. The top node is labeled `echo "from"` and the bottom node is labeled `echo "to"`.
- The bottom status bar indicates the execution status: `✓ 0s completed at 10:52`.

**Right Screenshot (Execution View):**

- The browser address bar shows the URL: `810qorfo8id-496f72e9c6d22116-8050-colab.googleusercontent.com`.
- At the top, there are buttons: `Load`, `Save`, `Remove`, `Run`, and `Swap`.
- Below the buttons, there is a text input field containing `echo "from"`.
- The main area displays the same GUI diagram as the left screenshot, with the top node labeled `echo "from"` and the bottom node labeled `echo "to"`.
- The bottom status bar shows the execution time: `workforce: 2024-05-28 22:52:28.426510`.

- Continuous workflows for data recording

# Final thoughts

- Workforce is a pipeline designer for downstream analysis
- It runs processes in parallel and across servers
- Much more testing is required!