

A FAIR Approach to Scientific Data Analysis with Boutiques

Gregory Kiar, Erin Benderoff, Jérôme Charrière, Simon Dubé, Louis-Olivier Guérin, Axel Bonnet, Sorina Camarasu-Pop, Valerie Hayot-Sasson, Stuart Thiel, Natacha Beck, Pierre Rioux, Serge Boroday, Shawn T. Brown, Jean-Baptiste Poline, Alan C. Evans, Tristan Glatard

What does it mean for a tool to be FAIR [1]?

Learn more at boutiques.github.io !

Findable

- Globally persistent records
- Described with rich metadata
- Searchable

We leverage **Zenodo** [2] to create DOIs for Boutiques descriptors which can be accessed via the Zenodo API.

Accessible

- Easily retrievable
- Universal access
- Persistent metadata beyond data lifetime

The retrievable tool descriptions contain **immutable** human- and machine-readable instructions for testing and launching each tool.

Interoperable

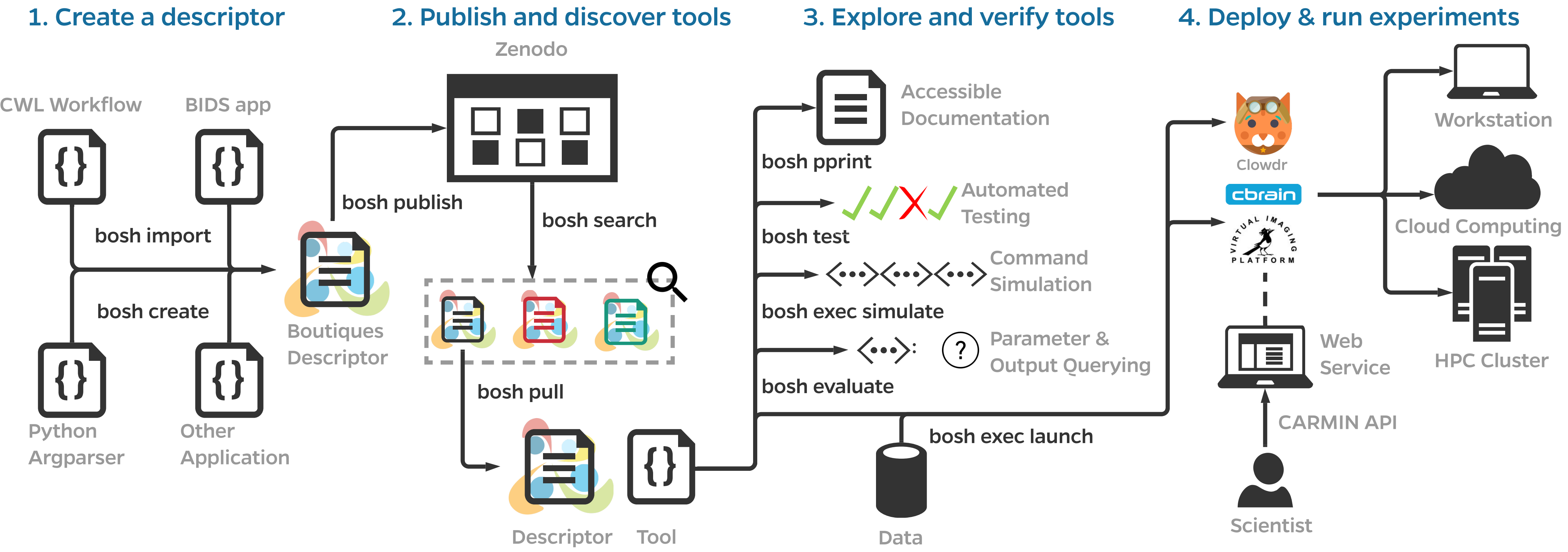
- Formalized and shared metadata standard
- Metadata standards adopted are FAIR
- Linking between objects where appropriate

CARMIN [3] and **Boutiques** [4] standards are used to describe and launch tools, either locally or through a RESTful API.

Re-Usable

- Multiple accurate and relevant attributes
- Clearly licensed
- Meets minimum domain standards

Docker [5] and **Singularity** [6] virtualization enable re-runability across platforms and enclosed testing. Simulation and querying allow runtime evaluation.



Using Boutiques

```
# Install the Boutiques package
$ pip install --user boutiques

# Use the shell utility (bosh) to search published tools
$ bosh search
ID          TITLE          DESCRIPTION          DOWNLOADS
zenodo.1895219 BIDS App - fmrprep fMRIPrep is a functional magneticresonan... 15
zenodo.2587156 blastdbcheck       BLAST database integrity and validity ch... 14
zenodo.1450983 PostFreeSurferPipelineBatch-CentOS7 PostFreeSurferPipelineBatch HCP pipeline 10
zenodo.1484547 BIDS App - FreeSurfer 6.0 BIDS App version of freesurfer 6.0, from... 10
zenodo.1450997 FreeSurferPipelineBatch-CentOS7 FreeSurferPipelineBatch HCP pipeline 9
zenodo.1494312 fsl_first         FSLT is a model-based segmentation and ... 8
zenodo.2566455 BIDS App - FSL Diffusion Preprocessing Preprocessing pipeline for diffusion MRI... 6
zenodo.2541125 BEST             <p>EEG/MEG source localisation technique... 4
zenodo.2566450 BIDS App - ndmg   ndmg connectome estimation pipeline 3

# Simulate parameters for a tool
$ bosh example zenodo.1494312
{
  "affine": "f.affine.96.mnc",
  "brain_extracted": true,
  "input_file": "f_input_file.32.tex",
  "method": "str_method.Zp",
  "prefix": "str_prefix_3a",
  "verbose": true
}

# Test the execution of a tool
$ bosh test zenodo.1472823
Exit Code: 0

# Use tools on your data
$ bosh exec launch zenodo.1494312 inputs.json
Shell Command
exampletool.py -c ./config.txt -i 'fo '""'; echo FAIL' bar -s 'coin;plop' -e vail ./setup.py
'log-4-coin;plop.txt' -l 1 2 3
Container location
Local copy
Container command
docker run --entrypoint=/bin/sh --rm -e ENVAR='theValue' -v
/Users/greg/code/gkiar/boutiques/tools/python:/Users/greg/code/gkiar/boutiques/tools/python -w
/Users/greg/code/gkiar/boutiques/tools/python -e HOME=$PWD boutiques/example1:test
/Users/greg/code/gkiar/boutiques/tools/python/temp-391546532120-1552931144164.localExec.boshjob.sh
Exit code
0
Std out
This is stdout
Std err
This is stderr
Error message
Output files
log-4-coin;plop.txt (logfile, Required)
- ./config.txt (config_file, Required)
Missing files
-
```

Boutiques utilities and the FAIR principle they facilitate

	Description	FAIR Principle
Import	In the case of existing applications described in the Common Workflow Language [7] or that are compliant with the BIDS application standard [8], the importer can automatically create a Boutiques descriptor.	I
Create	In cases not covered through importing, Boutiques can either create descriptors automatically from Python argparse-described tools or a blank template which can be populated.	I
Test	Adding tests to descriptors allows running an analysis in a predefined setting so that performance can be evaluated against expected results.	R
Publish	Boutiques leverages Zenodo to provide a free and permanent service which allows for storage and indexing of resources, and provides a unique DOI for each entry.	F
Search	Published descriptors can be discovered programmatically, including keyword filtering such as tool name, modality, author, or others.	F
Pprint	A pretty-print function creates documentation from a descriptor which is easily human readable and can be used to recreate a tool's command-line.	A
Simulate	Simulations can either be generated from random inputs or accept user-provided inputs to mock execution the of the tool.	R
Evaluate	Provided input fields and produced output files can be evaluated from an execution, allowing the automated identification of produced results.	R
Launch	Finally, the launch method allows for running tools with provided inputs.	R

References

[1]: M. D. Wilkinson et al., Sci Data, vol. 3, p. 160018, Mar. 2016.
[2]: Zenodo, <https://zenodo.org>
[3]: T. Glatard et al., Front. Neurosci., vol. 9, 2015.
[4]: T. Glatard et al., Gigascience, vol. 7, no. 5, May 2018.
[5]: C. Anderson et al., IEEE Software 32.3 (2015): 102-c3.
[6]: G.M. Kurtzer et al., PloS one 12.5 (2017): e0177459.
[7]: P. Amstutz et al., Mar. 2016.
[8]: K. J. Gorgolewski , PLoS CB, vol. 13, no. 3, p. e1005209, Mar. 2017.
[9]: G. Kiar et al., Front. Neuroinf. vol 13, 2019
[10]:T. Glatard et al., Zenodo, 10.5281/zenodo.2574166

