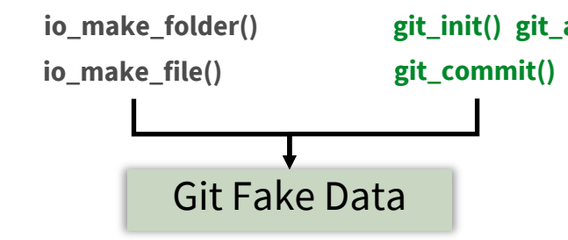
Fake Data Generator : : **CHEAT SHEET**

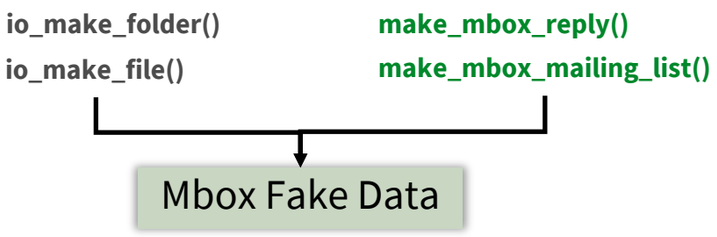
About

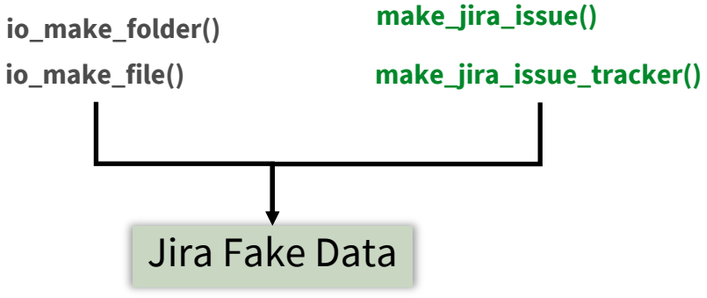
The fake data generators provide API to create minimal reproducible examples (MRE) capturing unusual project data to unit test Kaiaulu data parsers. They can also be used to compare equivalent analyses to other tools, and different versions of the same tool.

Fake Data Generator (git.R, mail.R, jira.R): APIs for Git, Mailing list, and Jira formats
Fake Examples (example.R): Minimal reproducible examples (MRE) generated on the fly building on Fake Generators
Unit Tests (test-*.R): Tests use MREs to test the behavior of functions

Fake Data Generator







Fake Examples

Writing Fake Examples

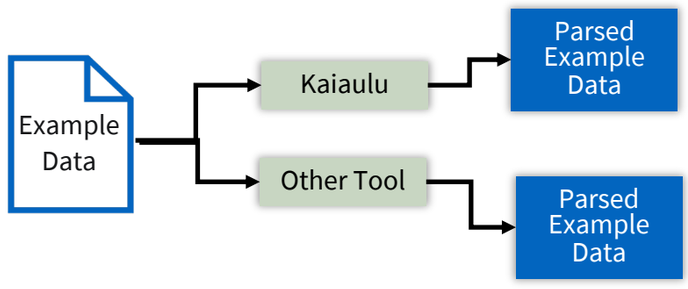
```
example_dif_branches <- function(...){

  file_path <- io_make_file(...,body="print('Hello!')")
  git_init(...)
  git_add(file_path,...)
  git_commit(...)
  ...
 return(example_path)
}
```

Fake examples use the Fake Data Generator API to annotate (and version) specific details in datasets that may be overlooked in analysis. They can then be passed to other functions in Kaiaulu:

```
parse_gitlog(path=example_dif_branches(),
tool="Perceval")
parse_gitlog(path= example_file_rename(),
tool="Perceval")
```

Comparing Tools Outputs



Fake Data Examples can also be used to compare Kaiaulu output to other tools and papers.

Often, research publications will also include examples to illustrate a method. These can be expressed in Kaiaulu as example functions.

Unit Testing

Unit Testing Fake Examples

```
test_that("Only Master Branch Commits are parsed",{
   git_repo_path <- example_dif_branches(...)
   expect_equal(parse_gitlog(git_repo_path),2)
}
```

Unit tests can be combined with fake examples to evaluate parser functions, which require data to be evaluated.

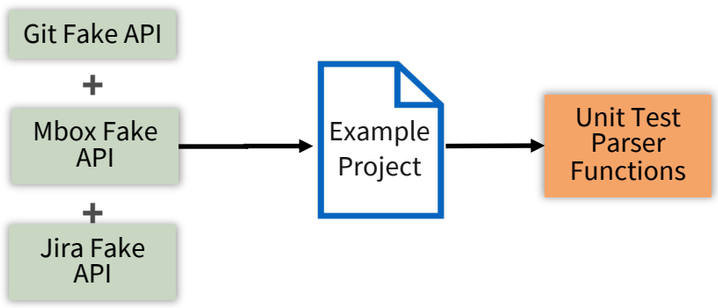
Third-Party Tool Behavior "Contract"

```
parse_gitlog(example_two_branches, tool="Perceval 1.0")
parse_gitlog(example_two_branches, tool="Perceval 2.0")
```

Some of Kaiaulu's functionality is borrowed from other tools. While new releases may introduce a change in analysis behavior, examples can be used to ensure output of interest remains consistent across versions.

Example datasets can also be used to compare different tools output.

Fake Ecosystems and Integration Testing



When combined, Fake Data Generators can be used to simulate a moment of interest in a project ecosystem. Analyses that utilize multiple sources can be sanity-checked alongside their code behavior.