



Reproducible
Analytical
Pipeline

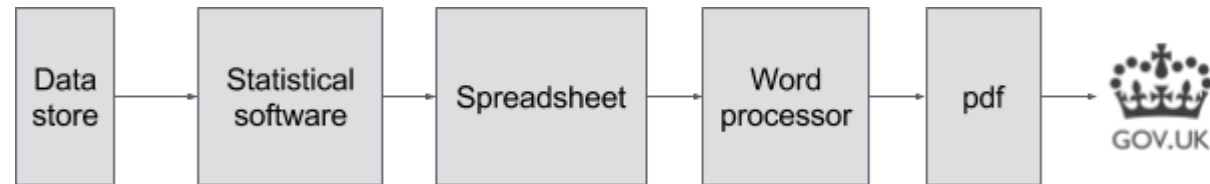
For Maps!

What is RAP?

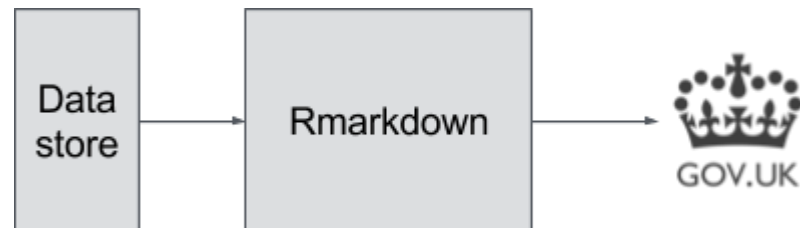
- RAP stands for *Reproducible Analytical Pipeline* and as the name suggests the aim is to create **reproducible** code to streamline a repetitive process such as a statistical processing system.
- RAP is used in some areas of UK government to update publication documents every time new data comes in to avoid the manual copy and pasting of data from one place to another.

What is RAP?

Conventional statistical workflow



Reproducible Analytical Pipeline



Why RAP?

I'm not happy because the statistics I want are not published 😞. I have to ask the department to produce the statistics I want which will take ages! ⌚⌚⌚⌚.

I'm unhappy because I'm repeating the same menial process over and over again 😞😞😞. I don't feel that I add value to the work...the skills that I learned at university are wasted.

I'm not happy because the statistics take weeks to produce, and my neck is on the line if there are errors 😞. My team is frustrated with the work, but we still need to get it done on time.

Person

Analyst

Manager

Stress-o-meter: 😞

😞😞

😞😞😞😞

*Made up for demonstration purposes

Udemy

Why RAP?

- - **Less error prone**
- - **Quicker**
- - **Reproducible - anyone can run it and get the same output!**
- - **Built in tests allow better quality assurance**
- Traditionally, QA has been a manual process which can take up a significant portion of the overall production time of a publication, as any changes will require the manual process of production to be repeated.
- - **Less repetition**
- - **Develop programming skills**

Main RAP concepts

1. Version control

- If working with a few people on the project, version control is a good way to track progress and share code.
- It's also a good way of me sharing code that I've made with you. Github and Gitlab are good platforms for versioning and sharing code.



Main RAP concepts

2. Reproducible code

• 2.1 Data

- A folder where all of the data is stored that needs to be fed into the processing system.

2.2 Config files

- A file that hold information that is likely to change between time periods. For example: variables that store what the current month or year is. It is the only file that would need to be edited before a new publication.

```
current_month = 'January'  
current_year = '2018'
```

```
data_location = 'Users/emily/data'
```

Main RAP concepts

2.3 Functions

- Code that can be used to repeat a process many times e.g.

```
def sum_two_numbers(a, b):  
    return a + b
```

3. Testing

- Writing tests is like having a built in quality control to test that our functions are doing what we want them to do.

```
import pytest
```

```
# content of test_sample.py
```

```
def sum_two_numbers(a, b):  
    return a + b
```

```
def test_answer():  
    assert sum_two_numbers(3, 2) == 5
```