# Introduction

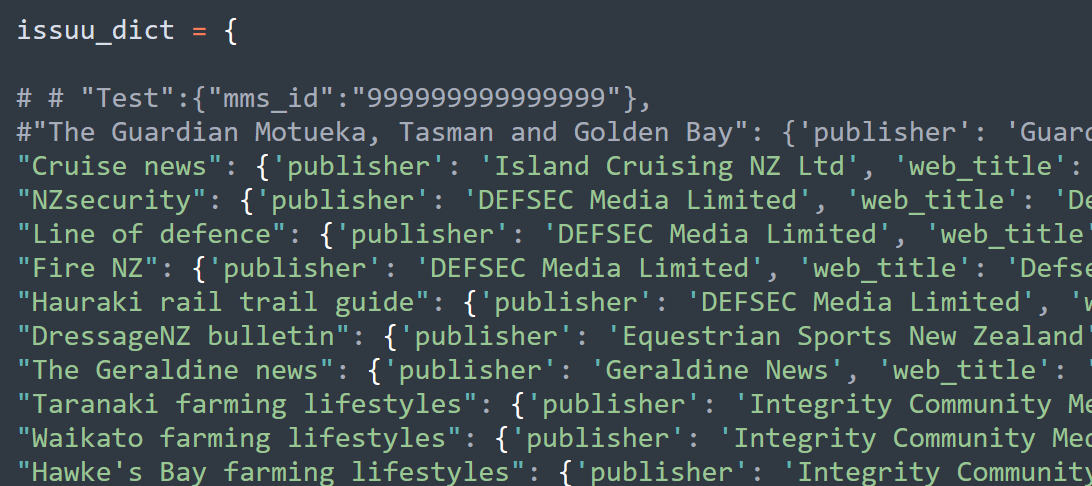
Currently there is to conjoined workflows. For collecting pdf and for images depending on “Download” option availability.

In the future this scripts could be joined for more comfortable work. Currently there is 3 main points to make title working dictionaries, reports and script rules. Changes should be added to each of this points .

# Dictionaries

issuu\_dict.py and issuu\_image\_dict.py

They contain initial serial and scraping information for each title and also starting point for harvesting. To activate new title - add it to dictionary. To exclude title for particular harvest it is enough just comment out the line with this title from dictionary.



Line could be made manually in the following format.

"Cruise news": {'publisher': 'Island Cruising NZ Ltd', 'web\_title': 'IslandCruising', 'mms\_id': '9919228892902836', 'holding\_id': '22383868940002836', 'po\_line': 'POL-187765', 'pattern': 'yyyy mm', 'days': '30', 'url': 'https://issuu.com/islandcruising', 'username': 'islandcruising', 'access': '200'} ,

Or if several titles come to the workflow it is comfortable to use **issuu\_populate\_dictionary.py**

Script contains 2 main functions which both read information from google spreadsheet and complete necessary Alma requests to populate other pieces of information such as pattern and holdings. Information which printed as output could be inserted then directly to one of the dictionaries for images or for pdf.

If you have google API credentials you can use “google\_spreadsheet\_routine()” function, otherwise download current version of spreadsheet as Excel spreadsheet and use “routine” function.

# Reports.

Second important part are reports as they are actively use to skip worked out titles and to make cover\_checker.html files.

**If title is new and only one magazine of the publisher on ISSUU this step could be skipped.**

Otherwise.

Reports are also located in 2 folders – reports and reports\_images. And for each title there 3 reports. Title.txt, (contains designation processed), **title\_worked\_out.txt** (contains ISSUU platform docnames processed), **title\_worked\_out2.txt** (contains docnames of other titles of the same publishers)

Reports are being populated during the process, however, due to huge amount of documents on ISSUU platform and also the titles might have prehistory and should be collected from certain time these workout reports should be created from the beginning.

Then script using them as exclusion, otherwise it should process all docnames of particular publisher.

To make the report process more comfortable use script **making\_files\_helper.py.**

After new title lines inserted to the dictionary other titles should be temporally commented out before running the script. If script is used for images workflow titles – do not forget uncomment 2 lines in the beginning of the script.

If the title is new, delete everything for title the **\_worked\_out.txt** file. And randomly in **\_worked\_out2.txt** title if publisher has more then one magazine.

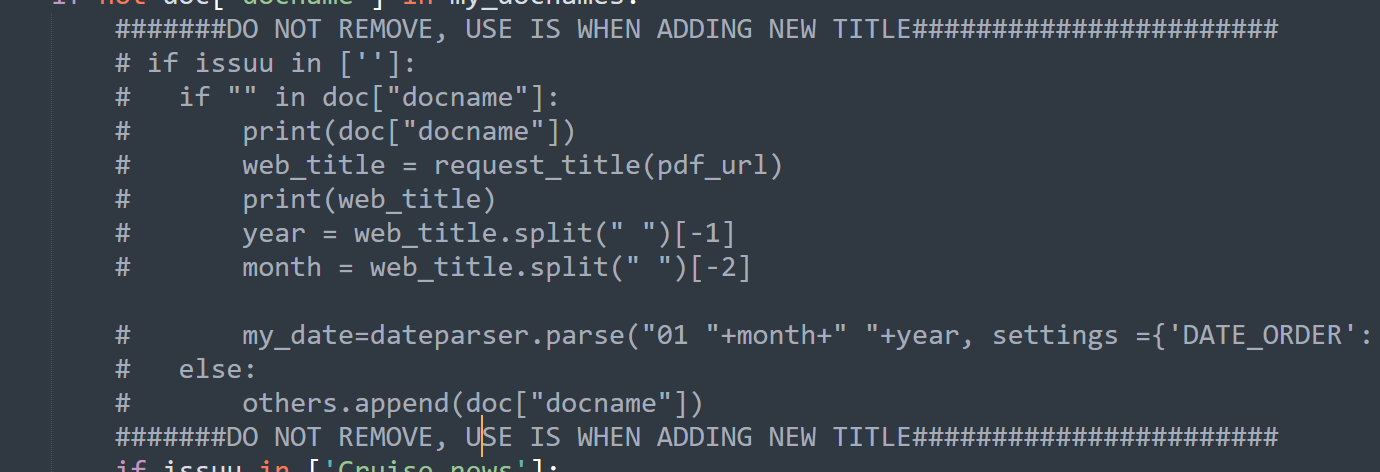
If some materials exist already, delete randomly from both files.

# Script rules

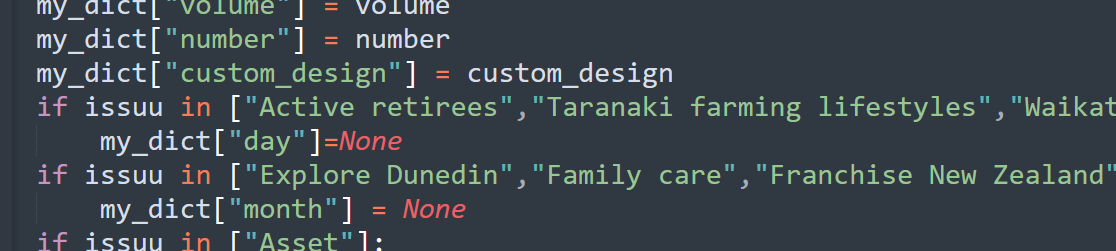
New title will not work until rules are not added to the scripts. - **issuu.py** or **issuu\_images.py c**orrespondently.

1. Copy existing template (it is in each script) to start making rules. Adjust them looking at existing rules for other publication. Year and month could be extracted from web\_title, from docname or from publication date on website(which is not much reliable). If there is a season use seas\_dict to find middle month to parse date.

Script decide if to collect title based on rules, existing in report issues and date, comparing it to last date existing in Alma for this title.

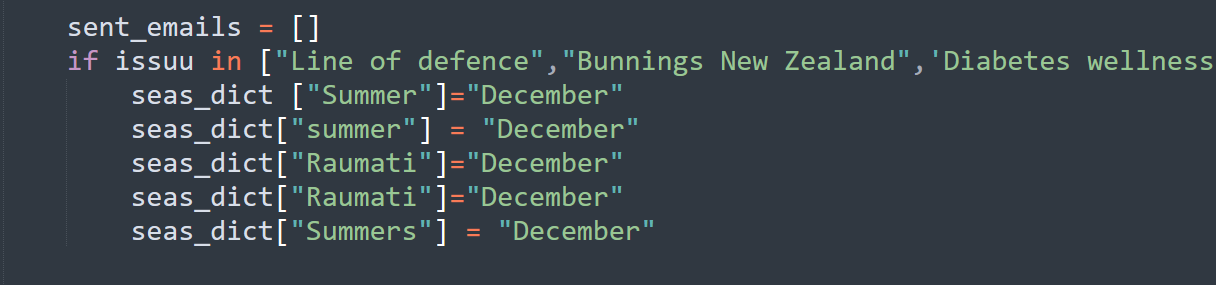


1. If necessary, add titles to the following lists.



* 1. If there is no day set it as 01 , add the title to the the first list in loop, wich set day as None back after date parsing.
  2. If there is not month set it as 01 as well and add title to the second list, to reset it back to None after parsing

1. In case if you publisher use season and they publish Summer issue as December, add the title to other list. It will help to avoid weird designation between two years.



# INNZ title

If title belongs to INNZ workflow, add it to the list in the end of **issuu\_dict.py**

innz\_titles = ["Asset", "At the bar","Diabetes wellness", "Education Gazette", "Family care", "Food New Zealand", "Forest and bird", "Human resources", "Navy today Royal New Zealand Navy", "New Zealand winegrower official journal", "Te karaka", "Te korowai o Tangaroa", "Heritage New Zealand", "NZ manufacturer success through innovation", "Pipiwharauroa", "Playmarket annual", "Real Estate", "SLANZA collected"]

This title goes then to the **innz\_titles\_innz.html**

**Done!**