Project Update – 29/07/2021

# To do’s

* Complete panel ranked list data collection (code development, testing, and running)
* Prepare person data for gender prediction.
* Start reviewing and cleaning the data.
* Finalise database schema.
* Start organising data by the schema.
* Search and collect data from Grants on the Web on grants and people.
* Investigate and create a table of linked grants (either similar cluster projects or continued projects)

# What you did and created?

* Implemented filtering functionality for finding valid links to Ranked Grant Lists
* Created the scrapping functionality for panel ranked grants lists from Grants on the Web:
  + Resolved issues that cropped when doing this like:
    - Only the first link being logged as being scrapped
    - Tables other than #dgGrants not being detected
  + Collected the Ranked grant lists, then cleaned, organised, and collated them into a single grants dataframe – still requires optimization and adding into the database schema
* Started cleaning the panel data and collating it into less, more organised dataframes:
  + Noticed some issues in the data collected caused by errors in the code – corrected and reran the data collection.
* Improved code commenting in Gateway to Research code

# What decision have you made and how they were made?

* Continued using sweet and wrapper functions to handle scrapping Panel ranked grant lists

# Relevant References

|  |  |  |
| --- | --- | --- |
| **Name** | **Link** | **Notes** |
| Advanced R | http://adv-r.had.co.nz/Exceptions-Debugging.html#condition-handling | Used to develop error handling functionality |
| Stack Overflow: Advanced TryCatch statements | <https://stackoverflow.com/questions/12193779/how-to-write-trycatch-in-r> | Used to develop complex error handling functionality |
| GtR-1-API-v3.3 | <https://gtr.ukri.org/resources/GtR-1-API-v3.3.pdf> | Used to develop API access expressions |
| Progress Package Readme | https://github.com/r-lib/progress#readme | Used vinettes/tutorial to develop progress bar functionality |
| Rvest | https://rvest.tidyverse.org/ | Used to develop scaping from HTML documents functionality |
| JSONLITE | <https://cran.r-project.org/web/packages/jsonlite/vignettes/json-aaquickstart.html> | Used to request and convert JSON files into dataframes |
| Subset list based on a filter | https://stackoverflow.com/questions/6941506/subset-elements-in-a-list-based-on-a-logical-condition | Used to filter the elements of a list based on a condition and explain the use of the Filter function |
| rvest Error in open.connection(x, “rb”) | https://stackoverflow.com/questions/33295686/rvest-error-in-open-connectionx-rb-timeout-was-reached | Used to solve error with html document retreval and develop more stable method. |
| [How to convert variable (object) name into String](https://stackoverflow.com/questions/14577412/how-to-convert-variable-object-name-into-string) | https://stackoverflow.com/questions/14577412/how-to-convert-variable-object-name-into-string | Used in logging functionality to return the variable name parsed into a function |
| Regex101 | https://regex101.com/ | Used to develop regex expressions |
| DrawSQL | <https://drawsql.app/> | Used to create database schema |
| *Nas Introduced by Coercion* | [*https://statisticsglobe.com/warning-message-nas-introduced-by-coercion-in-r*](https://statisticsglobe.com/warning-message-nas-introduced-by-coercion-in-r) | *Used to solve an issue when coercing data to the correct data type where lots of warnings would be thrown.* |
| *Remove periods/dots in entire data frame* | [*https://stackoverflow.com/questions/55026965/remove-periods-dots-in-entire-data-frame*](https://stackoverflow.com/questions/55026965/remove-periods-dots-in-entire-data-frame) | *Used to develop method to remove thousand seperators from values in dataframes* |
| *Change column order* | [*https://dplyr.tidyverse.org/reference/relocate.html*](https://dplyr.tidyverse.org/reference/relocate.html) | *Used to have a more elegant method of reordering columns than using a select* |

# To do items for coming weeks

* Review and clean the grant data
* Finalise database schema, correcting issues and typos.
* Start organising data by the schema
* Create a basic RShiny app
* Carry out two-three basic analysis based on the already collected data.

## Deferred To Dos:

* Prepare person data for gender prediction.
* Continue reviewing and cleaning the panel data.
* Investigate and create a table of linked grants (either similar cluster projects or continued projects)
* Search and collect data from Grants on the Web on grants and people.