### **ETL (Group-5 :** WellWell Project)

**Team members:** Ray Camo and Narisara Kantanong

**Project description:**

The outline of this project is to consolidate oil and gas well data from different government websites to the database. This data aggregation enrich and improve accessibility of the data as a single source database to the wider range of end-users.

More than 8000 well activities can be found from an individual Australian government website. Each state and territory in Australia also have their own database focusing on various data types and controls.

By combing data which existing in one sector but not another would reduce time searching for those data and enhance possibility of working on those dataset.

We limited our scope down to 3 wells as a pilot project where the dataset were cleaned, sorted, filtered, combined and transferred to the database.

From Bonaparte Basin, well Plover 1, Plover 2 and Whimbrel 1 data will be used in this project. With time constraints, only a few number of attributes mentioned in this project were also taken into consideration.

**Project activity:**

Data mining

For well headers information, wells were searched and downloaded using basin filter called “Bonaparte Basin”. Well headers information exported as nopims.xlsx.

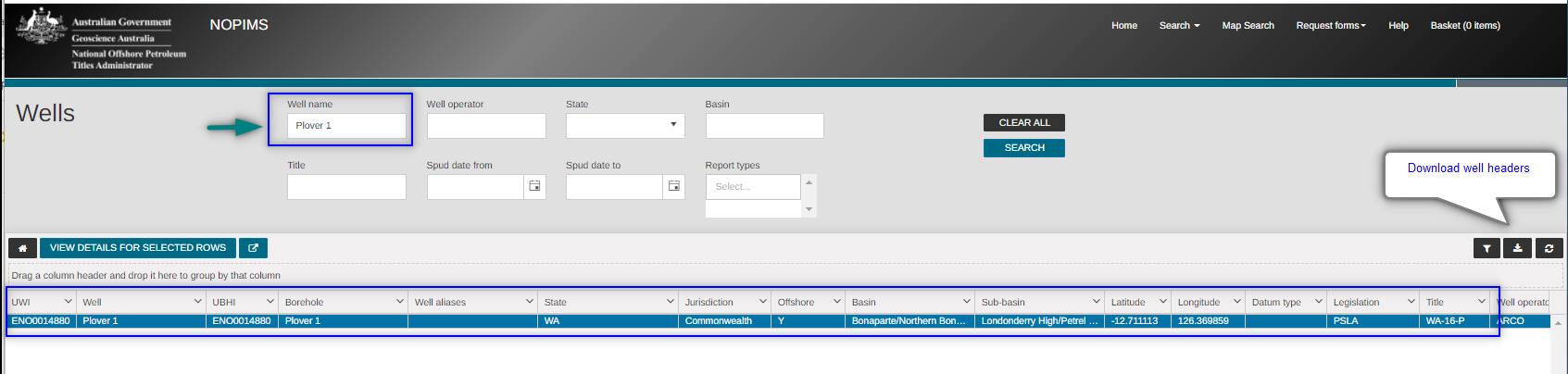
For well documents, we manually searched by well name (Plover 1, Plover 2 and Whimbrel 1) then exported list into excel files (Plover1.xlsx, Plover2.xlsx and Whimbrel1.xlsx).

*Source 1 - National Offshore Petroleum Information Management System (NOPIMS)*

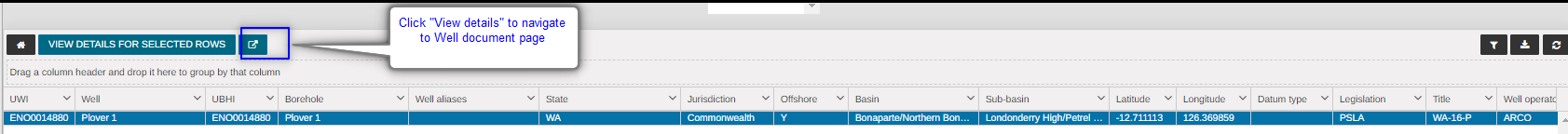
*Exported: Well headers information, Well documents list*

*Data format: .xlsx*

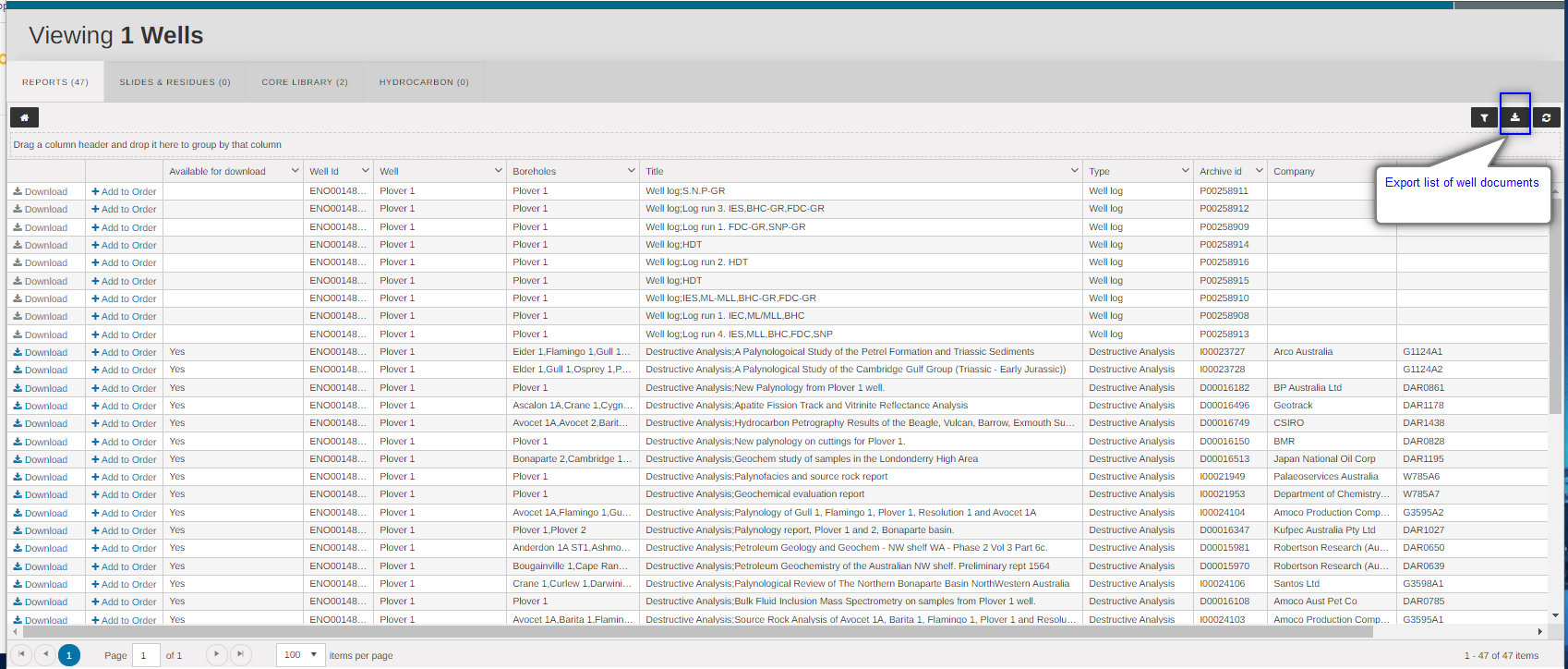
1. Navigate to well, for example below insert “Plover 1” under “Well name” then select download to retrieve well headers data.



1. Select well “Plover1” and select “View detail” to navigate to well documents



1. Download list of Well documents

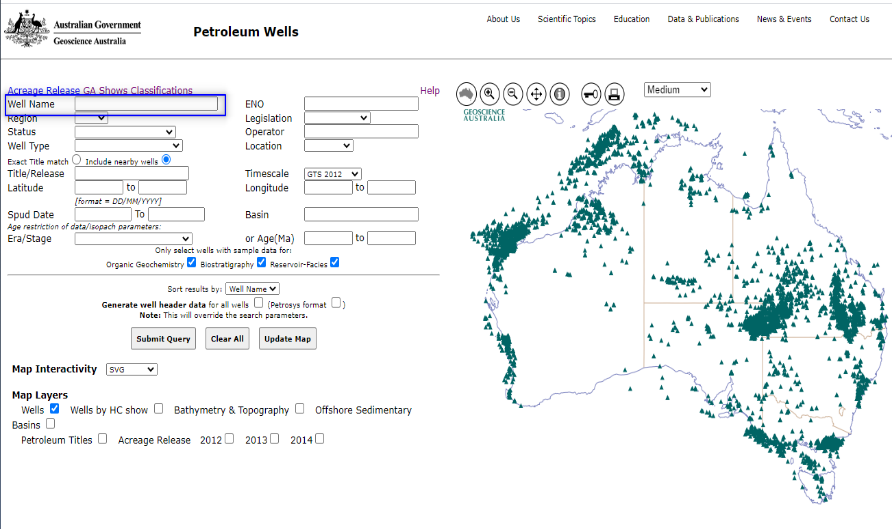


Source 2– Geoscience Australia

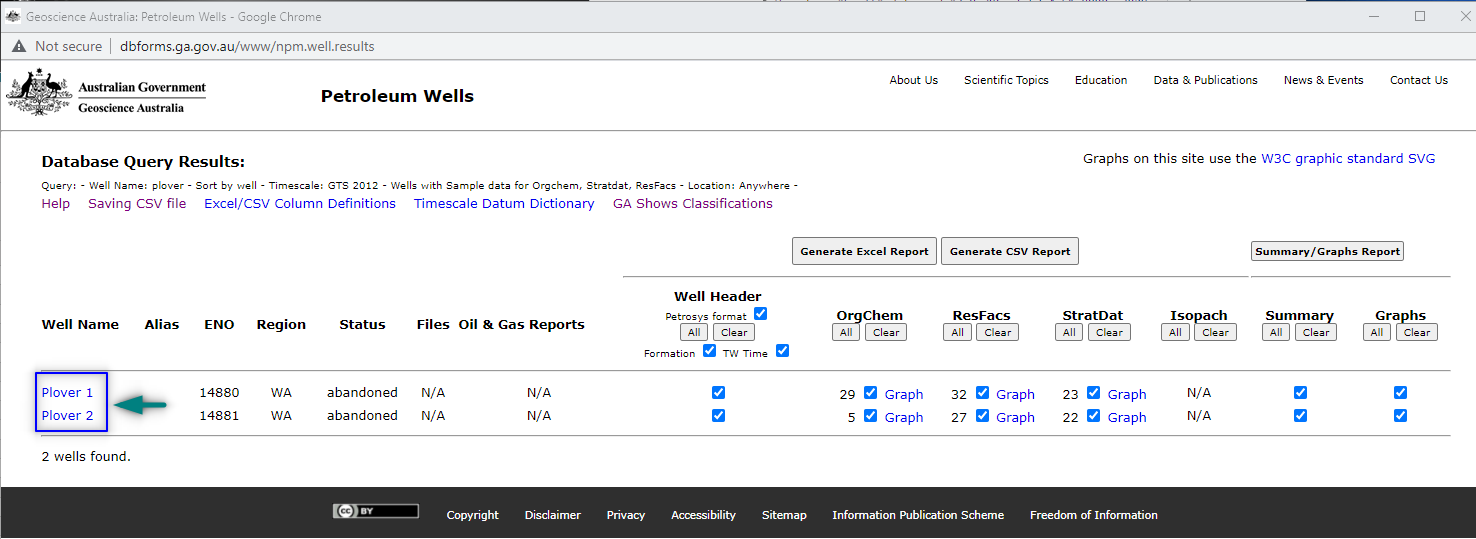
Exported: Summary Report

Data format: .html

1. Navigate to Geoscience Australia, then insert well name into the query table, for example “plover”



1. List of wells will show as below after submitted query. Selected all data available then clicked “Plover 1” to view and export Summary Report



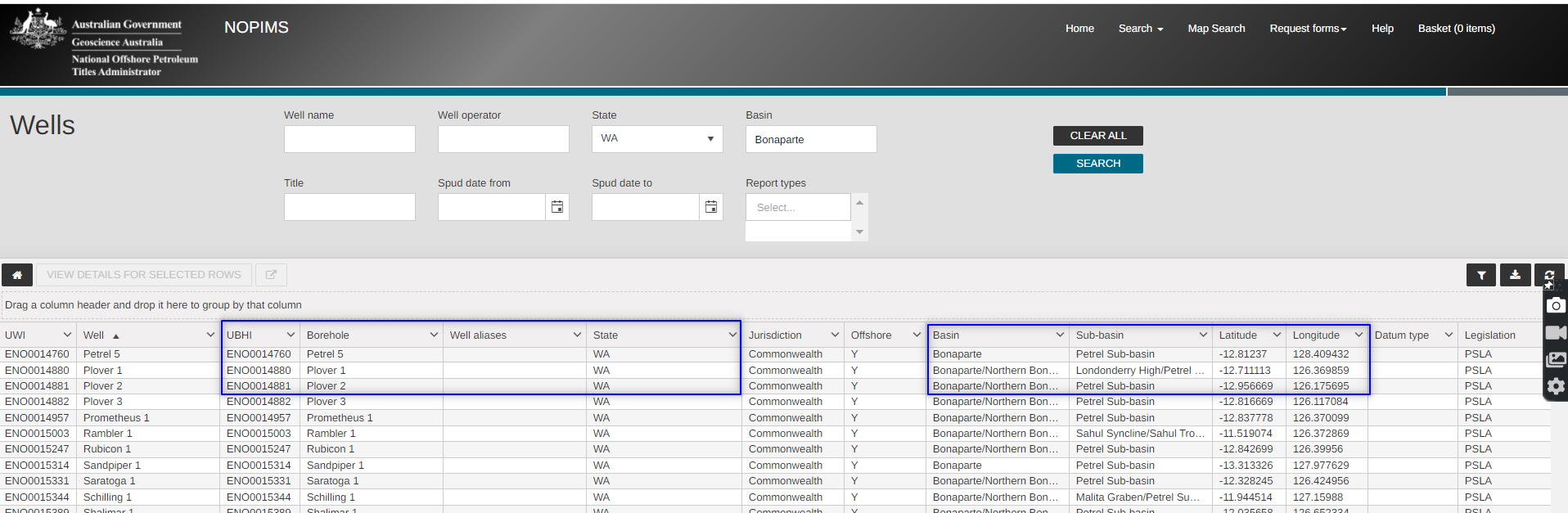
**Data cleaned up**

*Source 1 – NOPIMS*

*Data format: .xlsx*

Well names and well header attributes such as Borehole name, locations, and basins are filtered from nopims.xlsx into DataFrame called well\_header and basin. Unique ID also assigned to well Plover 1, Plover 2 and Whimbrel 1 as 1, 2 and 3 respectively.

Well documents exported from NOPIMS were sorted, the unique borehole\_id then added, and combined into the DataFrame.



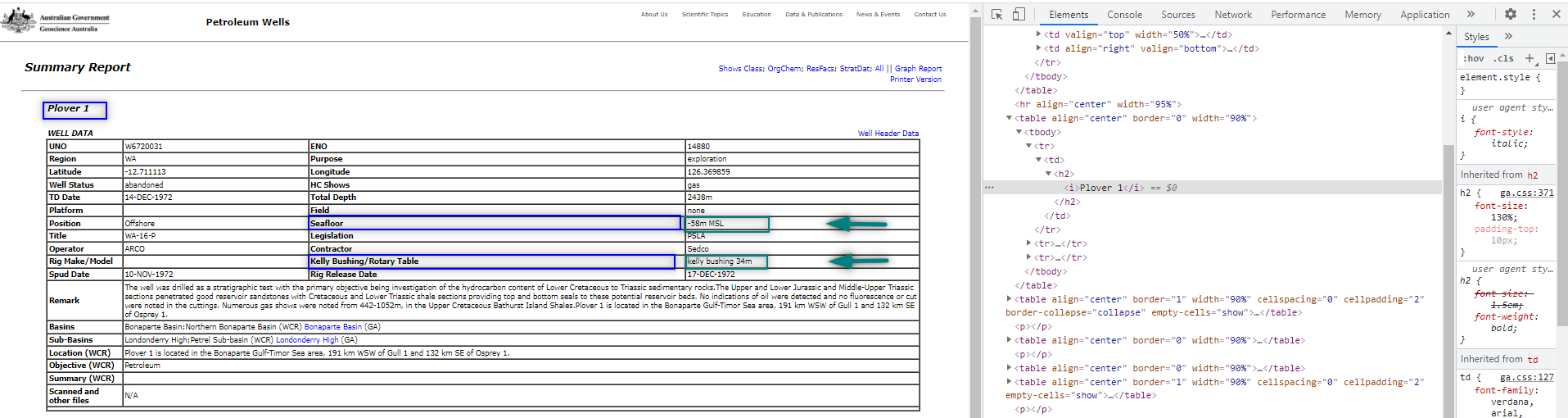
*Source 1 – GA*

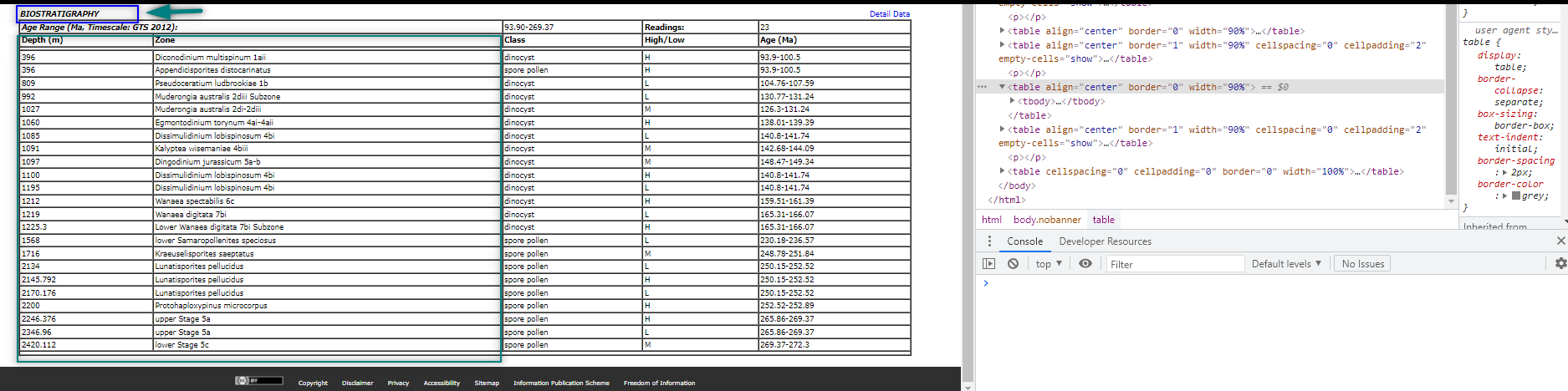
*Data format: .html*

Elevation and sea floor data were extracted from .html files using pandas. The unique borehole\_id then was revaluated to the combined DataFrame called evaluation\_combine.

Biostratigraphy data were also extracted from .html files using pandas, the unique borehole\_id was added to the DataFrame and then combined data into biostrat\_combined.

Example : Seafloor depth, Well elevation and Biostratigraphy data extracted from .html files (Summary Report)

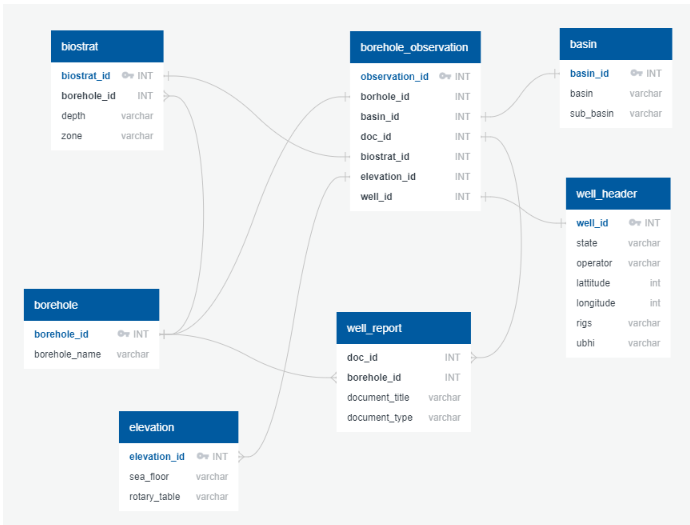




**Data analysis and Load**

**Schema Diagram**

Schema diagrams were drawn to create relationship between tables and data attributes.



Borehole\_observation table was added to the schema linking all attributes.

Each final DataFrame imported to table in PostgreSQL then transferred into wellwell\_db database.

**Summary**

Final dataset loaded to database can also be added, edited, deleted and updated, in the future. It can hold more customised information to suite specific group of end-users compared to what available in an individual government database.

An application such as spot fire or power bi can also be built on top for easily visualizing and data retrieving in the future.

Future improvement when working on large number of wells, web scraping could possibly be done during data mining stage to capture metadata directly form website.

**Data Sources and Tools:**

[https://nopims.dmp.wa.gov.au/Nopims/Search/Wells#](https://nopims.dmp.wa.gov.au/Nopims/Search/Wells)

http://dbforms.ga.gov.au/www/npm.well.search

QuickDBD: https://www.quickdatabasediagrams.com/