Fluid Data Description

Contents

[1 Introduction 2](#_Toc90629484)

[2 Overview 3](#_Toc90629485)

[2.1 Data Collection 3](#_Toc90629486)

[2.2 Data Extraction 3](#_Toc90629487)

[2.3 Data Access 3](#_Toc90629488)

[3 Fluid Data Database 4](#_Toc90629489)

# Introduction

Fluid Data is in the process of developing an application that streamlines access to open file data and reports provided by the government in the oil and gas industry. Anyone has free access to the data and reports stored on the government websites, however, much of the data is poorly organised and difficult to access and the most valuable information is buried inside the report files, usually in PDF format.

Fluid Data intend to systematically reorganise the available data to make it easier to find and process as well as extracting as the additional data from the reports themselves and making it usable and accessible.

# Overview

The main functions of the Fluid Data application are:

1. **Data Collection**: Collecting metadata and files from the state government database, cleaning and rearranging the data and then storing it in the Fluid Data database.
2. **Data Extraction**: Extracting additional data from report documents and storing them in the Fluid Data database
3. **Data Access**: Providing easy access to the data through a functional web-based application and through a simple Application Programming Interface (API)

## Data Collection

Open file well data must be collected from each Australian state separately. While the goal of Fluid Data is the combine the data from all Australian states and territories into a single location, Queensland has been the focus to date.

The Queensland geoscience portal has an API that can be used to access metadata on wells and reports in the Queensland database, including direct links to files and reports so they can be easily collected and analysed.

The Queensland data is provided under a created commons license allowing anyone to:

* Share - copy and redistribute the material in any medium or format
* Adapt - remix, transform, and build upon the material for any purpose, even commercially.

But must give appropriate credit, provide a link to the license, and indicate if changes were made.

## Data Extraction

Using machine learning to report files from the government database are converted to a text array format. The textual representation of the data will be stored in the Fluid Data database to enable further data collection and for users to easily search for text in the documents.

Pattern recognition is used to analyse the text arrays and extract additional data that would normally be difficult to discover and process.

## Data Access

Access to the data will be primarily through the Fluid Data web application. The web application will provide easy ways to search the database to extract the data and download files and reports.

Users will also be given access to an Application Programming Interface (API) which they can use as an alternative to the web application interface, or can be integrated into their own software packages to link the Fluid Data database with their existing data.

# Fluid Data Database

The class diagram on the following page demonstrates the structure of the database. The main tables of the database are:

* The Well table, consisting of:
  + gov\_id: the unique identifier of the well in the government database
  + modified: the date that data was last updated in the government database
  + well\_name: The name of the well
  + latitude and longitude: geographical location of the well
  + rig\_release: the data the drilling rig was released from site
  + operator: the company currently operating the well
  + state: the state the well is located in
  + permit: the licensing permit the well is operated under
  + well\_status: the current usage status of the well
  + well\_class: see notes on class diagram
  + well\_purpose: the resource that is targeted by the well
* The Document table, consisting of:
  + document\_name: the title of the document
  + url: the link the document on the government server
  + well: the id of the well the document is associated with
  + file: details of the local file on the Fluid Data server
* The Page table identifies each page of a document after the page has been extracted and saved on the server as a separate jpeg file. And addition data extracted from that page image is linked to the page in the database, allowing both the page image, the original document and the associated well to be quickly linked to the data.
* The Data table class provides the base of what will be a series of unique tables to store the various types of data that will be extracted. Some examples are shown in the class diagram (Desorption, Seams, DST). The extraction\_method field refers to the pattern recognition that was used to obtain the data.

