# **Portfolio B**

# OmniChem

# Overview:

In our daily life, data comes in many forms, such as: numbers, patterns or even symbols. Data are also used in many different fields to improve our economy, enhance technologies and research. However, in order to be able to use these data efficiently, a paramounted tool like data management – acquiring, storing, and processing data – is inevitable. Moreover, using data management to work with big data in research fields would the researcher task easier. For instance, acquiring a huge collection of data and store them in a database, where the researcher can simply view or export them into a readable and understandable format.

With the lack of easy access and user-friendly tool in working with big data and data management, suitable for a specific type of researching field, we planned to implement a software that answers our client’s need. As, our client is Dr. Ella Gale, a chemistry machine learning doctor at the University of Bristol. Let along with chemistry researchers at the University of Bristol and other institutes, pharmaceutical research companies and others with a stake hold in easily managing chemistry datasets. The software must be able to aid the research procedures, as well as, assisting in new development, which could bring the clients success and fame.

In this sense, our system is a Chemistry Dataset management and search tool with the intent of being used for passing, storing, and organizing chemistry data from the lab machine. Our system will also come along with the capability to easily data mine in searching for the existing Chemistry related data, for instance, chemical bonds or chemical formula. In addition, the system must have an intuitive and well-designed GUI system, in exporting unreadable machine format, for the non-computer literate Chemistry research community. Most importantly, for security and privacy of the users’ information and their data, we also want this database to be held securely by using a unique username and password to access not only remote systems but also local files.

Such that, we intend to build this core database and data management system using an SQL and Java, while the GUI will be built using HTML, CSS and JavaScript. This thoughts in implementing the system are attentively planned by taking the future development into consideration, where the programing language for implementation are widely used in the computation fields. Thus, once the system is launched and used, not only us, but people with programming background would be able to keep enhancing the software to the better version

Chart, waterfall chart

Description automatically generated

# Requirements:

## List of Identified System Stakeholders and Descriptions:

|  |  |  |
| --- | --- | --- |
| Dr. Ella M. Gale | - | Post-Doctoral researcher in machine learning for chemistry at the University of Bristol. As well as, the major client who proposed the problems and oversees the implementation of the program. As a client and researcher, she would like to have a software program that is capable of import, query, sort, store, and read complex machine format dataset format. As well as, export and display the data in an easily readable format on a user-friendly Graphic User Interface. |
|  |  |  |
| Pharmaceutical Researcher/ Companies | - | Pharmaceutical researchers or companies are those who discover, develop, and produce drugs for hospitals or public uses, with the aim to cure or inoculate patients with symptoms. As pharmaceutical companies/ researchers, they thrive with a way to accelerate research by making data easier to sort through and share; thereby, assisting research in producing better and innovative drugs for the public. |
|  |  |  |
| General Public | - | The end-user of the whole process, the one who will be the most beneficial from the result of the software that aids drug production. The general public will get to use the new drugs that could be produced faster or invented by looking at the new data trends. |
|  |  |  |
| Chemistry Researchers | - | The chemistry researchers not only develop new drugs, they also improve the quality and efficiency of existing drugs and their manufacture. As a chemistry researcher, both working with commercial companies or institutions, their main aim is to analyze chemicals and create new drugs that could be able to help more people. Thus, they want to be able to easily manipulate and log all the data they have. It would definitely improve efficiency in many aspects of the research. |
|  |  |  |
| Chemistry Departments & Lecturer | - | The Chemistry department and lecturers not only providing the students in their department with knowledge of chemistry, they also supervise and take part in research within the university. The chemistry department and lecturers want to be able to use the product in the labs, which will aid students in their education. Such as by showing them examples of how the molecules or structures of the chemicals look like. Likewise, demonstrating to the student what kind of data coming from the new synthesis bot may look like. Furthermore, due to COVID-19, the department and lecturer give their students more data mining projects, which could be assisted by the software. |
|  |  |  |
| Drug Developer | - | As a drug developer, they want to gain access to large datasets and ease of logging data, which may improve their current manufacturing process. This takes into account of designing, enhancing, and producing new drugs that can cure more people. |
|  |  |  |
| Students | - | Individuals who enroll in the learning institute, in order to acquire knowledge and grow in specific professions they chose. The students, as the stakeholders, can range from middle school students to post-graduate or even PhDs. As students, they hope they could be able to see the patterns in data and efficiently use the results in my school’s projects and labs. Simply to enhance each individual knowledge and skills. |
|  |  |  |
| Machine Learning Developers & Companies | - | Machine learning developers and companies simultaneously advancing algorithms with the technologies/machines that are capable of automatically adapting and learning through its experiences. Also, getting their hands on the complex data structures and memories that help the machine to be able to grow and help humanities and their life. Moreover, by dealing with machines and data, they want to have access to those large datasets of chemistry data, which can aid both chemistry fields and machine learning fields. |
|  |  |  |
| University of Bristol / Universities | - | As an institution which provides higher education, along with excel oneself in research, the University of Bristol is ranked 5th for School of Chemistry and ranked 10th for School of Computer Science, in all over the UK. As a university that achieves a high ranking in both faculties, the university thrives with products from projects and researches that are able to give them fame and honors. Thus, they plan to accelerate research and produce new ideas or inventions which bring fames to the university itself. Also, by improving the chemistry departments data handling capabilities, with the help of the computer science department, the university could achieve more goals. This can also help increase the number of quality researches and continue moving up the rank. |
|  |  |  |
| JGI (Jean Golding Institutes) | - | The central hub for data science and data-intensive research at University of Bristol. The institute connects multidisciplinary community of experts across the university itself and beyond (researchers, cities, industry partners, and international partners), through data, in order to tackle challenges in data sets. |
|  |  |  |
| Research Systems Engineers | - | A group of individuals who specialized in writing programs or software. They are capable of developing industrial level software that are able to aid and solve multiple real-world problems. Thus, as professional individuals, they would want to continue working with the software, after the project has been passed on, for extension and development. |
|  |  |  |

## User Stakeholders

|  |  |  |
| --- | --- | --- |
| Chemistry Researchers | 🡪 | As a chemistry researcher, I have to deal with multiple formats of data that I fetched from different tools in the lab. The complexity in analyzing, manipulating, and using these data also varies.   1. Importing Data Story: I want a software with a user-friendly GUI that is capable of handling multiple formats of data from each distinct tool. 2. Searching Data Story: I want the software that is able store and recognize both existing and new dataset being input (working as some sort of database). I also want the software to be able to display those datasets in a viewable format. 3. Exporting Data Story: I want a saw that is able to export the datasets in some general format that could be use by other general tools or people (PDF or CSV). |
|  |  |  |
| Machine Learning Developers | 🡪 | As a machine learning developer, I want the GUI and each aspect of the program to be user friendly. Moreover, I want the software to be able to develop by itself, in term of memories, as it acts like database which store the existing and new datasets.   1. Import Data Story: I want a software that is capable of handling big datasets that come in various format. The software must be able to detect and proceed without my need to select the file format. 2. Search Data Story: I want to access the database which store all the datasets and fetch query for the one I want. The database or the search function must also be a user-friendly one, where the non-developer or beginner could observe and use it without any difficulties. 3. Export Data Story: I want an export function which use advance technologies, more than just exporting as a file. This could come in form of sharing the file to cloud or send to other people via email. |
|  |  |  |
| Pharmaceutical Researchers | 🡪 | As a pharmaceutical researcher, I want to be able to keep developing new drugs and improve the existing drugs, with the aim to cure symptoms that appear in the real world. I also want a software that could help increase the efficiency of drugs production.   1. Import Data Story: I want a program that collects the datasets from tools and sorts these datasets into appropriate categories. 2. Search Data Story: I want a software that has easy to query functions. Moreover, displaying and manipulating these datasets with some functions. 3. Export Data Story: I want the aiding software to excel in sharing these datasets via multiple ways; for instance, able to upload to cloud or exportable in simple format. |

Table

Description automatically generated

## Flow Steps:

Chemistry Researcher –

|  |  |  |
| --- | --- | --- |
| Basic Flow | Alternative Flow | Exceptional Flow |
| Import Datasets   1. Login to the software 2. Select import bar 3. Import datasets file 4. View the datasets displayed   Query Datasets   1. Login to the software 2. Select search bar 3. Type the relevant keyword to the datasets 4. View information of datasets   Export Datasets   1. Login to the software 2. Select export bar 3. Select the datasets wanted to export 4. Choose the export file type 5. Export and Download | Import Datasets   1. Login to the software 2. Select import bar 3. Import datasets file 4. The datasets already exist! 5. View detail of the datasets   Query Datasets   1. Login to the software 2. Select search bar 3. Type datasets keyword 4. No datasets found 5. Import new datasets or cancel query   Export Datasets   1. Login to the software 2. Select export bar 3. Select datasets 4. Upload to cloud / share via email | Import Datasets   1. Login to the software 2. Error – Login failed, contact the IT team   Query Datasets   1. Login to the software 2. Error – Login failed, contact the IT team   Export Datasets   1. Login to the software 2. Error – Login failed, contact the IT team |

Machine Learning Developer –

|  |  |  |
| --- | --- | --- |
| Basic Flow | Alternative Flow | Exceptional Flow |
| Import Datasets   1. Login to the software 2. Select import bar 3. Import File (Not specifying the file type) 4. File is being extracted and read the string 5. String is transfer into machine language and store   Query Datasets   1. Login to the software 2. Select search bar 3. Search via keywords 4. Access database 5. Keywords are being query through local database 6. Found and use data to display with multiple advance features   Export Datasets   1. Login to the software 2. Select export bar 3. Select the exporting file format 4. Fetch datasets in the database 5. Transfer and print data on an exportable file format 6. Export as a file | Import Datasets   1. Login to software 2. Select import bar 3. Manually type in the datasets to import 4. Information are process into suitable format 5. Information are added to local memory and cloud   Query Datasets   1. Login to the software 2. Select search bar 3. Select data fields or category’s name on the navigation bar 4. The software query through database based on the categories being selected 5. Display all the dataset’s name inside each field selected   Export Datasets   1. Login to the software 2. Select export bar 3. Select the export file format 4. Fetch datasets in the database 5. Transfer and print data on an exportable file format 6. Share the file via email or cloud | Import Datasets   1. Login to the software 2. Select import bar 3. Import file 4. Error – Cannot connect to the local database 5. Wait for the reconnection   Query Datasets   1. Login to the software 2. Select query bar 3. Search with keyword 4. Error – Cannot connect to the local database 5. Wait for the reconnection   Export Datasets   1. Login to the software 2. Select export bar 3. Select the datasets and export format 4. Error – Cannot connect to the local database 5. Wait for the reconnection |

Pharmaceutical Researcher –

|  |  |  |
| --- | --- | --- |
| Basic Flow | Alternative Flow | Exceptional Flow |
| Import Datasets   1. Login to the software 2. Select import bar 3. Import datasets manually/file 4. View the datasets displayed   Query Datasets   1. Login to the software 2. Select search bar 3. Type the relevant keyword to the datasets 4. View information of datasets   Export Datasets   1. Login to the software 2. Select export bar 3. Select the datasets wanted to export 4. Choose the export file type | Import Datasets   1. Login to the software 2. Select import bar 3. Import datasets file 4. The datasets already exist! 5. View detail of the datasets   Query Datasets   1. Login to the software 2. Select search bar 3. Type datasets keyword 4. No datasets found 5. Import new datasets or cancel query   Export Datasets   1. Login to the software 2. Select export bar 3. Select datasets 4. Upload to cloud / share via email | Import Datasets   1. Login to the software 2. Select import bar 3. Import file 4. Error – Cannot import (File type is not readable) 5. Contact IT team   Query Datasets   1. Login to the software 2. Select search bar 3. Search datasets keyword 4. Select the dataset 5. Error – Cannot display the information 6. Contact IT team   Export Datasets   1. Login to the software 2. Select export bar 3. Select export to file and share 4. Error – the file cannot be shared 5. Contact IT team |

## Chemistry Researchers (User Story – Atomic Requirement):

Functional Requirements:

1. The researchers **must** be able to login to their account (or could create a new one)
2. The researchers **must** be able to import any type of file to the software (or input the datasets manually)
3. The researchers **must** be able to search for the datasets using any relevant keywords
4. The researchers **must** be able to scroll search for the datasets using navigation bar in search page
5. The researchers **must** be able to display the datasets in any viewable format (e.g. picture)
6. The researchers **must** be able to edit the datasets information
7. The researchers **must** be able to export the datasets in their selected file format
8. The researchers **must** be able to share the exported data to other people
9. The researchers **must** be able to report when there is an error occur

Non-Functional Requirements:

1. The search function **should** update as soon as new dataset has been imported into the database
2. The software **should** be easy to use by no technology background chemistry researcher (beginner or researchers who are not used to using this kind of software)
3. The software **should** not take a long time to import new datasets, search for the datasets or export the file.
4. The software **should** be able to open in multiple OS
5. The software **should** have a very large capacity (because it deals with large data)
6. The datasets **should** only belong to each account

Chart, waterfall chart

Description automatically generated

# Architecture:

Overview to approach

The basic premise of the project is that of a system that will allow chemists, to easily collate manage and search vast datasets.

With this in mind the core component of the system is a database that stores each data-set as a table. There then needs to be a database interface component that handles the manipulation of the database. On top of that is a user interface component that manages the interaction of the user [NFReq.2]. The system also needs to facilitate the conversion of various file types into the database so a data converter module will also be required [FReq.2][FReq.8]. We also need the capability for the system to access remote data stores on servers [FReq.9].

We chose for the system to run predominantly on a native app as the user will need constant access to their data (even through brown outs) [NFReq.2], much of the data is also sensitive in nature and so will need to be held locally where it is more secure. The users also expressed their want for the data to be held locally.

We chose for the system to run a local web app which can then be accessed on the local host via a browser. This will allow us to use spring for the system. It will also allow for the easy deployment of a remote server system as both the client and server will run the same code. This will also save us allot of development work

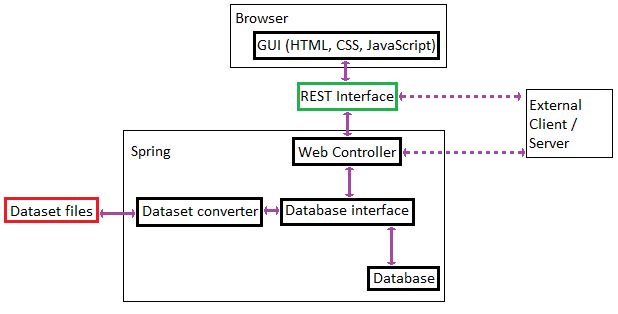
The system will need to scale to potentially thousands of users so has been designed to scale in modules with each organisation that uses it hosting their own collection of servers while each of their staff run the client on their local machines.

Use of technologies

The system will use Spring Boot for managing the database, GUI requests and web requests.

The database will be SQUL based and either contained within the Java Spring application itself or run separately in parallel but be initiated and closed by the application.

The front-end GUI will be HTML, CSS and JavaScript based giving the system universality and scalability. Without the system tied to a GUI it will also be far easier to test as tests can be performed through the REST API instead of the GUI.



Chart, waterfall chart

Description automatically generated

DEVELOPMENT TESTING

One key function of our system is that the clients can edit (import, export, remove or download) datasets to a server. It is basically the purpose of our whole system. One use of this function would be to allow the clients to manipulate and log data easier as all datasets will be shared in a common server to all the clients. Therefore, current manufacturing processes will be improved as the clients will have access to large datasets. For example, it would help pharmaceutical companies to accelerate research by making data easier to sort through and share.

|  |  |  |  |
| --- | --- | --- | --- |
| **Test scenario** | **Test case** | **Expected result** | |
| Check functionality of importing datasets to the server | 1.Check response on importing valid datasets  2.Check response on importing invalid file due to the dataset having invalid characters. | 1.New dataset <name> added  2.Invalid dataset/incorrect file type. | |
| Check functionality of downloading datasets from the server | 1.Check response on downloading a valid dataset from the server  2.Check response on downloading an invalid dataset from the server (invalid characters, or out of range values) | 1.<name> dataset downloaded  2.invalid dataset | |
| Check functionality of removing datasets from the server | 1.Check response on removing a valid dataset on the server  2.Check response on not being able to remove a dataset from the server as the dataset has already being removed. | 1.<name> dataset removed  2.<name> dataset cannot be removed. | |
| Check functionality of data mining from available datasets | 1.Check response on correctly exporting a dataset.  2.Check response on incorrectly exporting a dataset due to the dataset being semantically incorrect | 1.<dataset> exported  2.error <dataset> couldn’t be exported. | |
|  | | |

Release Testing

For high-level release testing, we will write a script which a team member acting as a user will follow. The tasks described in the user stories will make up most of the script as they should cover most aspects of the system, there may also be separate sections to cover specific parts of the system such as customisable settings. The team member will record their screen as they carry this out, audio will then be overlayed describing what is happening. This may also be accompanied by a text description with timestamps.

Core User story: Importing/creating a dataset (Chemistry researcher)

We’ve chosen this as it is a vital part of the system to stakeholders such as chemistry researchers and students, it also covers multiple important components in the system such as reading in data, checking inputted data and adding datasets to the database.

Tests for user story

|  |  |  |  |
| --- | --- | --- | --- |
| Action | Flow | Input | Expected result |
| Sign-in | Basic | Username then password | User is taken to sign-in page |
| Alternate | Password than username |
| Exceptional | Absent username and or password | User is told to fill in missing information |
| Incorrect username | User is informed login details are incorrect |
| Incorrect password |
| Click create button | Basic | NA | A popup prompts the user for details to create the dataset |
| Get data for dataset | Basic | User clicks local button and selects TestData file | File is selected as data source for Dataset |
| User clicks manual and types correct chemistry data | Entered information is used as data source for Dataset |
| Alternate | User clicks manual entry, enters some data, clicks okay then clicks manual entry and adds to it before hitting okay again | Entered information is used as data source for Dataset |
| Exceptional | User clicks local button and closes the popup | User is taken back to previous page however any information entered is maintained |
| User clicks manual and closes the popup |
| User clicks local and selects an invalid file (e.g. an image) | An error message informs the user that their file is invalid, the file explorer stays open |
| User clicks manual and tries to proceed without entering information | An error message informs the user that they can’t create an empty Dataset, stays on manual entry page |
| User clicks manual and enters invalid chemistry data | An error message informs the user what is wrong with the entered information, stays on manual entry page |
| User clicks manual, enters information and tries to close the popup | A warning message asks the user if they are sure they want to close the window as they will lose the entered information |

Note that actual data isn’t in the inputs as they will be large datasets.