Data Digitalisation in Medicine Manufacturing

Tabbasum Naz, Antony Vassileiou, Blair Johnston

tabbasum.naz@strath.ac.uk, antony.vassileiou@strath.ac.uk, blair.johnston@strath.ac.uk

Continuous Manufacturing and Advanced Crystallisation (CMAC), Strathclyde Institute of Pharmacy & Biomedical Sciences, University of Strathclyde, Technology and Innovation Centre, 99 George Street, Glasgow G1 1RD, UK.

1- Introduction

Data are the fuel for modern Industrial Digital Technologies (IDTs) such as AI. If data are not FAIRified then it could take many years for the pharma sector to reach a position in which the full value of IDTs are realised. Beyond the slow pace of data curation, the high level of laborious human intervention, and associated costs, will continue to be a challenge. Through the Smarter Innovation Centre - Digital Medicines Manufacturing Research Centre (DM^2) project funded by EPSRC, we provide a foundation on which we can begin to build trusted and structured data sets that will significantly improve reusability and future value. Within Platform 1 of DM^2, we have developed Extract-Transform-Load (ETL) tools to simplify data acquisition efforts and allow future data to be integrated easily. The DM^2 ETL tool, with multiple components, has been developed for automatic extraction, transformation and loading of heterogeneous medicine manufacturing data from multiple instruments. Schema for experimental data in the medicine manufacturing domain has been designed that provides a structure for data and establishes linkage to meta-data. Once data from multiple instruments is structured through DM^2, data can be visualised by a powerful data visualisation and business intelligence software Tableau. This helps the domain experts in medicine manufacturing to see and understand their data. It provides an intuitive and interactive way to explore, analyze, and visualise data from various sources. It allows users to create visualisations without the need for extensive programming skills. Medicine manufacturing data can also be accessed by domain experts in Artificial Intelligence for modelling.

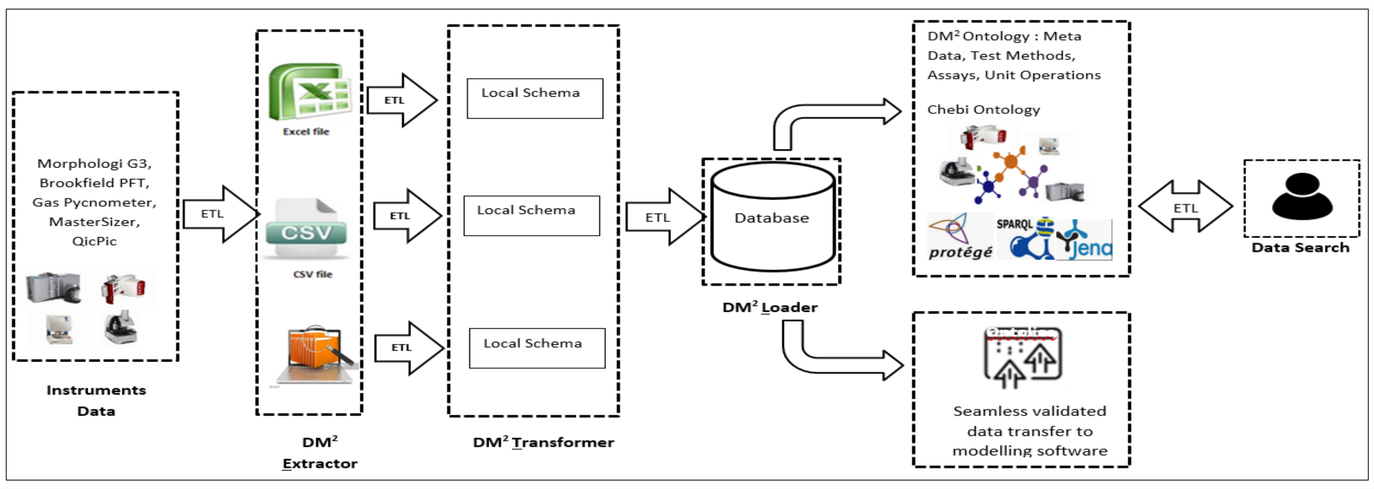
2- Methodology

ETL Extractor: The Extractor component is responsible for extraction of schema/data from different raw data sources i.e., Morphologi G3 and Gas Pycnometer. Data from equipment are currently not machine readable and are available in multiple heterogeneous formats making it difficult to integrate. The Extractor component resolves the problem of format heterogeneity by accessing multiple data formats.

ETL Transformer : The Transformer component provides automatic techniques for schema/data transformation and resolves heterogeneity issues.

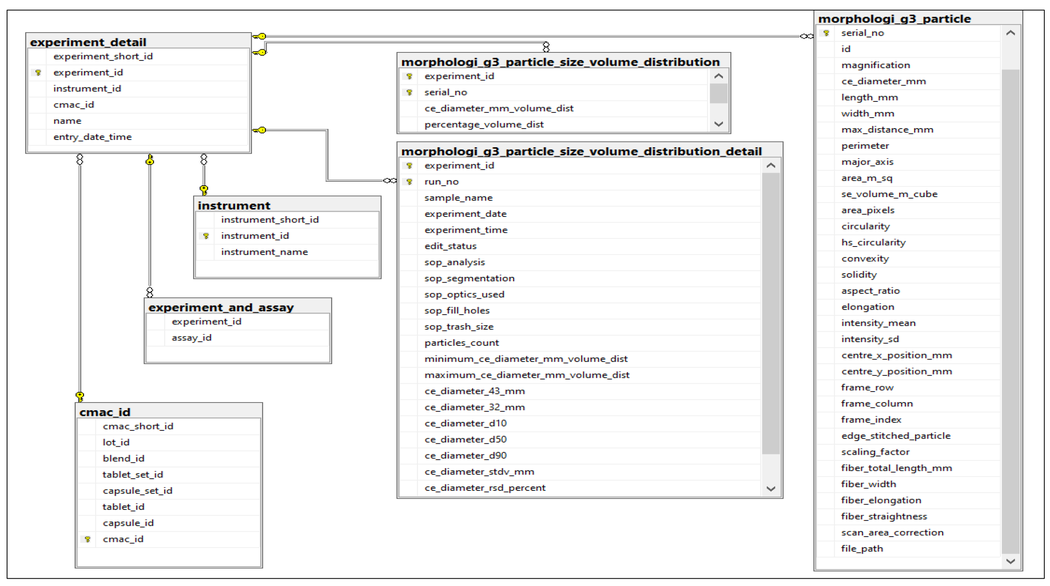
ETL Loader: Finally, the Loader component is responsible for loading the data into the DM^2 repository that provides access to structured data. With the help of DM^2 structured data, the content is readily available to AI tools for future research.

3- Software Architecture



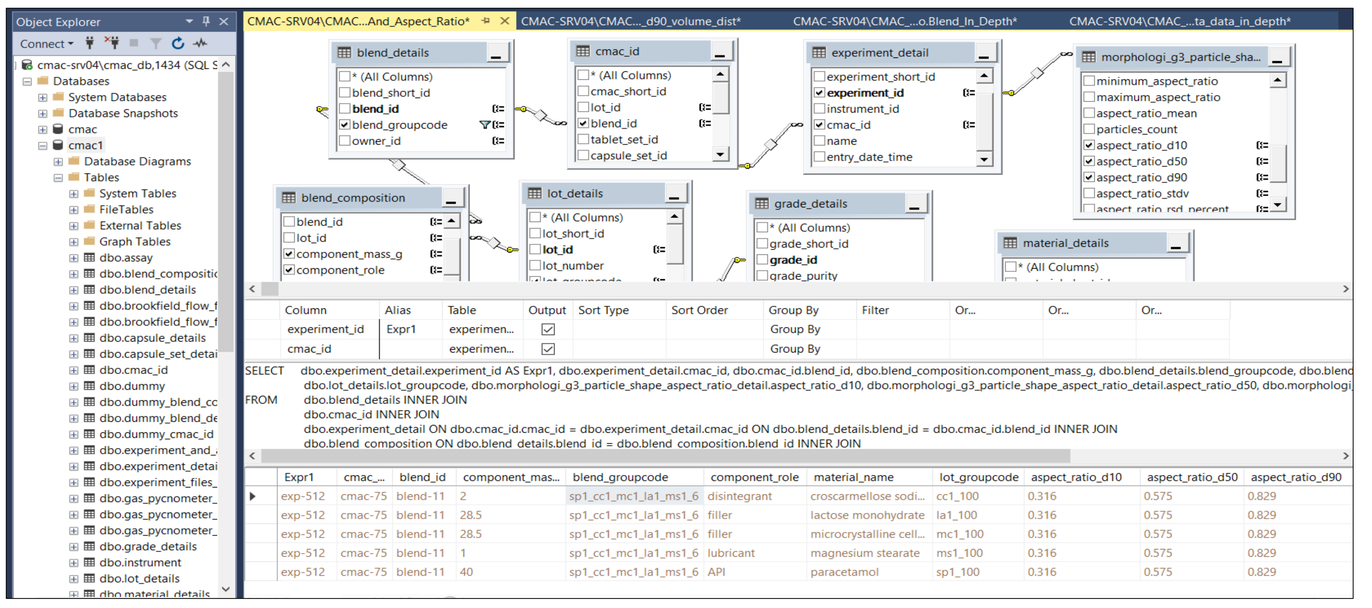
4- Experimental Data Modelling

- Entity Relation diagram below shows data modelling for experimental meta data, particle and volume distribution for Morphologi G3.
- In the similar way, schema for particle size number distribution, particle shape aspect ratio for Morphologi G3 and gas pycnometer has been designed.
- Linkage of instrument's data with meta-data.



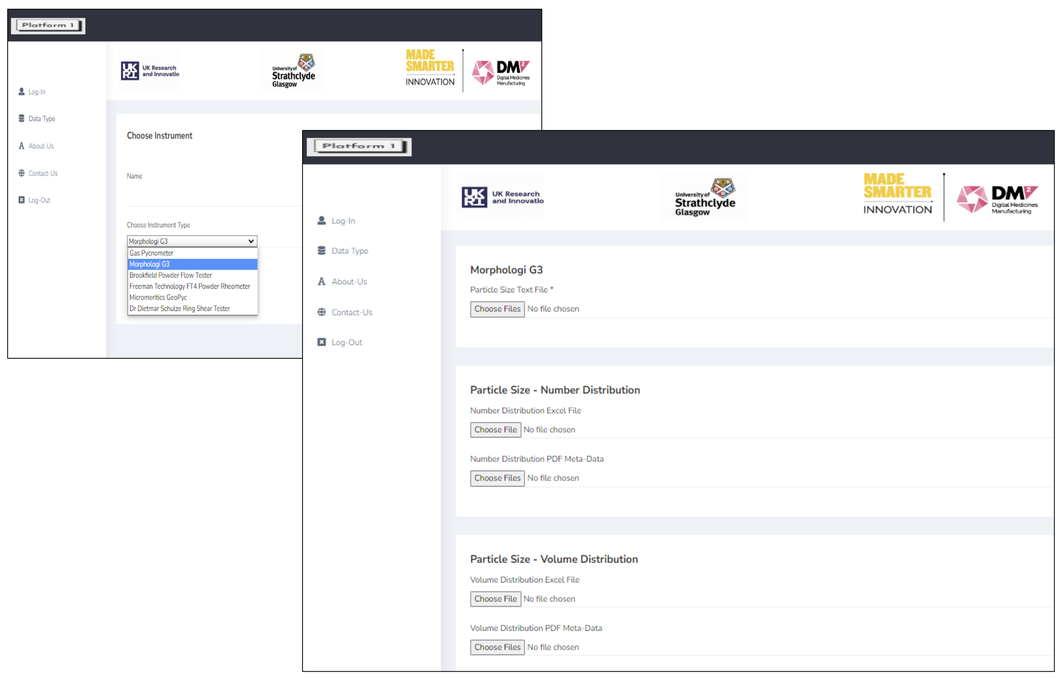
6- Search Data

- Search blend related experiments.
- Extract experiments with blend group code, components mass, component role and aspect ratio (d10, d50 & 90) for Morphologi G3.



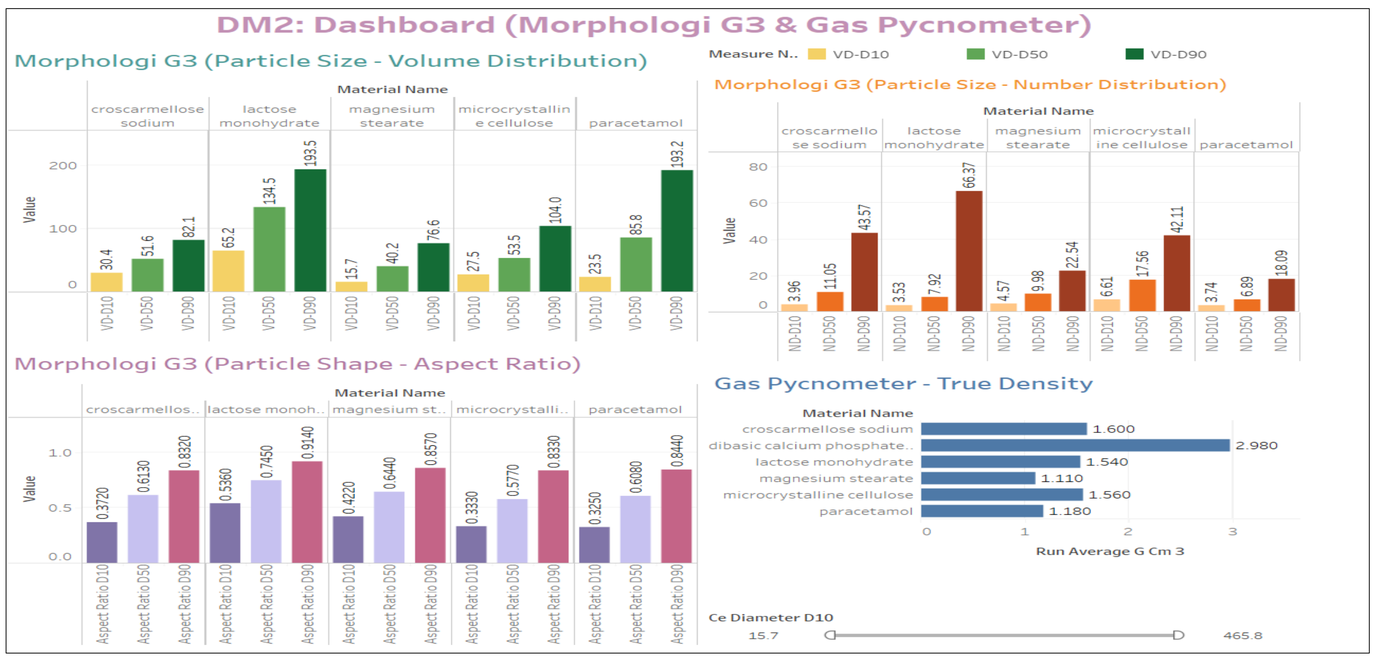
5- DM^2 ETL (Extract Transform & Load) Tool

- Interface to automatically extract, transform and load instrument's data via DM^2 ETL
- Data generated by platform 2 for DM^2 project has been loaded.



7- DM^2 - Data Visualisation

- Data visualisation of i) particle size volume distribution, ii) particle size number distribution, iii) particle shape aspect ratio from Morphologi G3 and iv) true density from Gas Pycnometer for different materials in a dashboard by linking data from DM^2 central repository to Tableau visualiser.



8- Conclusion and Future Work

We have developed DM^2 ETL (Extract, Transform and Load) – a tool that can extract medicine manufacturing data from different sources and develop a mechanism to translate between different concepts and data from multiple schemas. Data modeling and visualization of the Morphologi G3 and Gas Pycnometer instruments has been covered. In future, we plan to develop a semantic layer to data search with the help of domain ontology in the medicine manufacturing domain. Domain ontology for meta-data will be used as a way to represent meta-data. In future, data from DM^2 ETL can be reused by experts in AI, predictive analysis, statistical analysis, data visualization, data mining and machine learning.

Acknowledgment: The authors would like to thank the Digital Medicines Manufacturing (DM^2) Research Centre (Grant Ref: EP/V062077/1) for funding this work. DM^2 is co-funded by the Made Smarter Innovation challenge at UK Research and Innovation, and partner organisations from the medicines manufacturing sector. For more information, visit cmac.ac.uk/dm2-home

DM^2 website