

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 (DM²) 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 DM², we have developed Extract-Transform-Load (ETL) tools to simplify data acquisition efforts and allow future data to be integrated easily. The DM² 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 DM², 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 DM² repository that provides access to structured data. With the help of DM² structured data, the content is readily available to AI tools for future research.

4- Experimental Data Modelling

- \succ 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.



assay_id			particles_count		intensity_sd
			minimum_ce_diameter_mm_volume_dist		centre_x_position_mm
			maximum_ce_diameter_mm_volume_dist		centre_y_position_mm
			ce_diameter_43_mm		frame_row
smac id		ce_diameter_32_mm		frame_column	
CITIC	cmac short id		ce_diameter_d10		frame_index
	lot id		ce_diameter_d50		edge_stitched_particle
	blend id		ce_diameter_d90		scaling_factor
	tablet set id		ce_diameter_stdv_mm		fiber_total_length_mm
	capsule set id		ce_diameter_rsd_percent	~	fiber_width
	tablet id	I			fiber_elongation
	capsule id				fiber_straightness
	cmac id				scan_area_correction
•	cmac_ia				file_path

.dummy .dummy_blend_cc .dummy_blend_de .dummy_cmac_id .experiment_and_a	FROM	dbo.lot_details.lot_groupcode, dbo.morphologi_g3_particle_shape_aspect_ratio_detail.aspect_ratio_d10, dbo.morphologi_g3_particle_shape_aspect_ratio_detail.aspect_ratio_d50, dbo.morphologi_d3_particle_shape_aspect_ratio_detail.aspect_ratio_d50, dbo.morphologi_d4bo.blend_details INNER JOIN dbo.cmac_id INNER JOIN dbo.experiment_detail ON dbo.cmac_id.cmac_id = dbo.experiment_detail.cmac_id ON dbo.blend_details.blend_id = dbo.cmac_id.blend_id INNER JOIN dbo.blend composition ON dbo.blend details.blend id = dbo.blend composition.blend id INNER JOIN										
experiment_detail		Expr1	cmac	blend_id	component_mas	blend_groupcode	component_role	material_name	lot_groupcode	aspect_ratio_d10	aspect_ratio_d50	aspect_ratio_d90
as_pycnometer_		exp-512	cmac-75	blend-11	2	sp1_cc1_mc1_la1_ms1_6	disintegrant	croscarmellose sodi	cc1_100	0.316	0.575	0.829
s_pycnometer_		exp-512	cmac-75	blend-11	28.5	sp1_cc1_mc1_la1_ms1_6	filler	lactose monohydrate	la1_100	0.316	0.575	0.829
s_pycnometer_		exp-512	cmac-75	blend-11	28.5	sp1_cc1_mc1_la1_ms1_6	filler	microcrystalline cell	mc1_100	0.316	0.575	0.829
ade_details		exp-512	cmac-75	blend-11	1	sp1_cc1_mc1_la1_ms1_6	lubricant	magnesium stearate	ms1_100	0.316	0.575	0.829
nstrument ot_details naterial_details		exp-512	cmac-75	blend-11	40	sp1_cc1_mc1_la1_ms1_6	API	paracetamol	sp1_100	0.316	0.575	0.829

5- DM² ETL (Extract Transform & Load) Tool

- > Interface to automatically extract, transform and load instrument's data via DM² ETL
- \blacktriangleright Data generated by platform 2 for DM² project has been loaded.



7- DM² - 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 DM² central repository to Tableau visualiser.





8- Conclusion and Future Work

We have developed DM² 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 DM² ETL can be reused by experts in AI, predictive analysis, statistical analysis, data visualization, data mining and machine learning.

DM² website

1.540







Acknowledgment: The authors would like to thank the Digital Medicines Manufacturing (DM²) Research Centre (Grant Ref: EP/V062077/1) for funding this work. DM² 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

lactose monohydrate

