



Olga López | Digital Technology Architect

Digital Platform Group

Data Architecture

Oct 24th, 2024



Agenda

- **Purpose:** To update on Data Architecture progress and governance rules.
 - Enterprise Data Architecture & Ent Data Model by Aneesha Soman
 - Data Factory evolution by Abhiram Gandhe
- **Approach:** open discussion and validation first by domain architects.
 - When EDM should be used?
 - How EDM/OSD is embedded into Data Factory capability and products?
- **Value:**
 - Adherence to DS Smith Data strategy establishing a standard procedure in Data Factory to inform projects of the value, implications and cost of EDM / ODS approach.
- **Audience:** Domain Architects



Summary Analysis on EDM/ODS/Dimensional & Tera example

The strategic choice to follow an enterprise data model at the heart of our data factory through to its build within key products has not been without debate, disruption and disconnect. CTO findings are:

- **Concept:** The concept of **EDM** alignment was established at July '24 **Enterprise Data Council** following on from the development of the EDM itself. Translation into **logical** and **physical** form was beyond the level of detail of EDC.
- **Logical:** The approach to build the EDM into data factory started with a **dimensional data model**.
 - Requirements from the emerging **Data Architecture Group** (CTO, Packaging) sought a physical instantiation through and **Operational Data Store**. This is documented in PPT and presented to the Data Factory Design Group **11/06/24**.
 - No common document on the drivers, goals and benefits of this.
 - No documented steps to assess and communicate any impact on existing data, inflight or planned projects.
- **Physical:** There is no mechanism (funding and resource) to proactively build the Data Factory ODS framework.
- **Tera Phase 1:** There is no record of a requirement from ARB, TDA or other for Tera 1 to follow the ODS.
- **Tera Phase 2:** Caught at August 2024 TDA requiring conformance to ODS. Request for re-estimate by supplier.
- **PSCT:** At October IT Steerco, PSCT is giving the option back to the business to choose with/without ODS.

01 Enterprise Data Architecture





Data Architecture

Why is this important?

If data is a new asset for DS Smith, just like any of our facility or plant assets, it is easier to manage in a consistent way if we are able to describe what it is, its hierarchy, components, capabilities, condition, quality and more. Data Architecture is a discipline as well as an output set of models that makes this possible. It is a prerequisite to data engineering, data preparation, data integration and data quality.

What is our strategy?

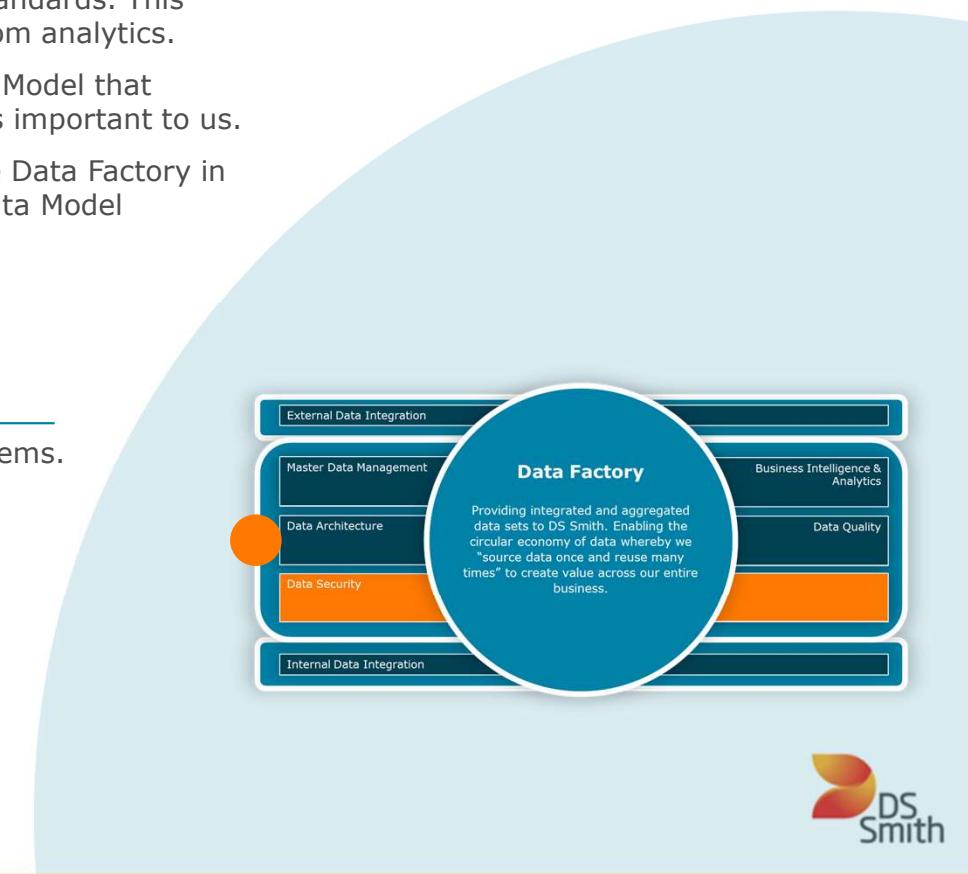
- We will document our data assets, mapping how data flows through our systems.
- We will map change initiatives to the data that they need in order to guide prioritisation and dependency management
- Build our data factory according to the structures of the Ent Data Model
- Build our Integration Hub according to the Ent Data Model
- Data architecture encompasses a wide range of components, including data models, data integration, data storage, and data governance.

What does it look it today?

- The data landscape is fragmented and of significant variety and standards. This erodes business value from analytics.
- Our first Enterprise Data Model that describes the data that is important to us.
- First steps in shaping the Data Factory in accordance to the Ent Data Model

Technology

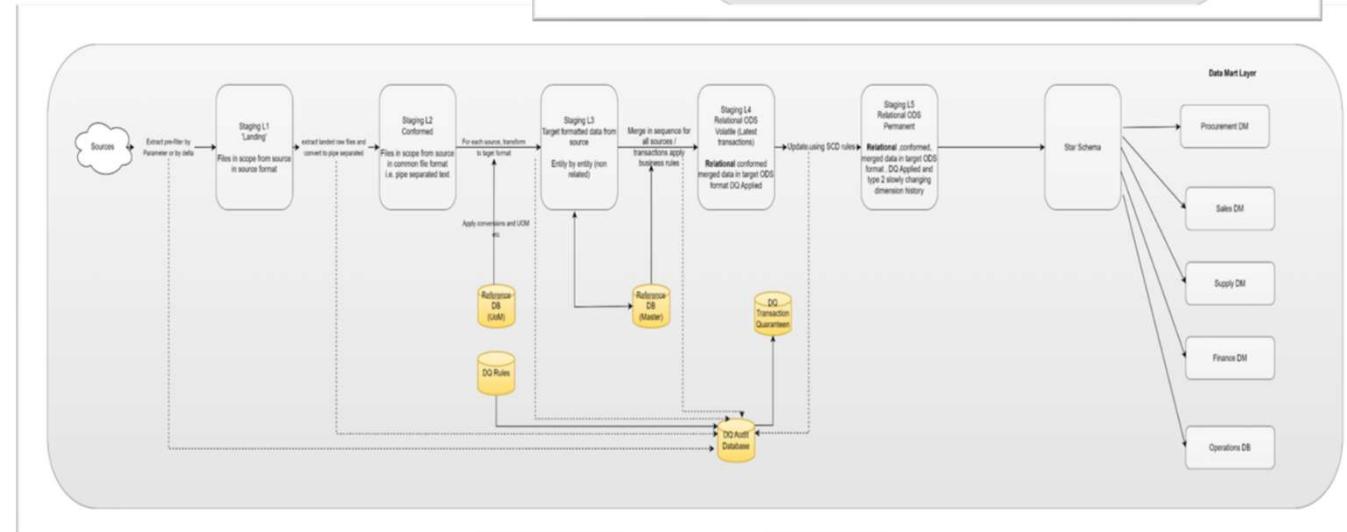
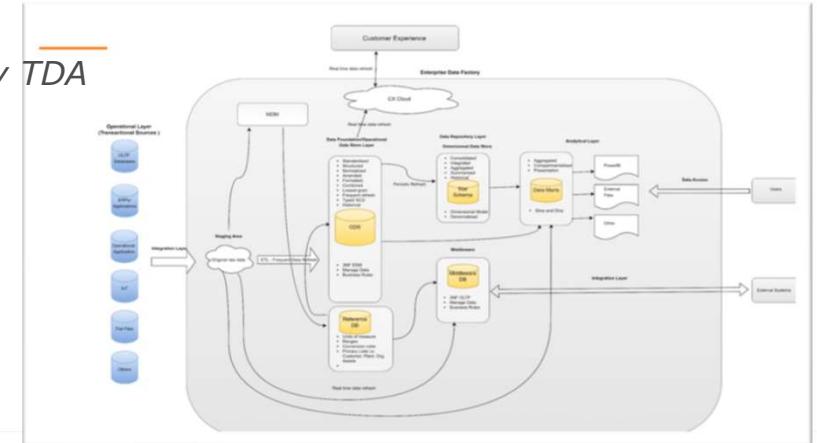
- Some basic data modelling tools



Enterprise Data Architecture

- EDM should exist in Data Factory in a Physicalised 3rd Normal Form.
 - Increase **reusability**, **eliminate redundant data**, ensure **data integrity** and **quality**.
- The understanding is that the initial cost will be high to implement the ODS but will decrease over time.
- **Full Agreement** on call that the EDM is a pattern that is mandated for delivery. (Agreed in “EDM latest position” in meeting last week)

**Not necessarily approved* by TDA*



Ent Data Architecture – Low level Details

Staging L1 – Landing : To land extracted data , filtered by parameter or delta , in their source format.

Staging L2 - Conformed : To Extract from Landing to conform into a common format .

Staging L3 – Target Format :

1. To transform entity by entity to target ODS format. At this stage data is non – relational
2. Applies conversion rules based on Reference DB (Unit of Measure)
- 3.DQ Rules are applied.
4. Master Data Update.

Staging L4 – Relational ODS :

1. To store conformed and merged data in sequence from L3.
- 2.Applies business rules on transactional data.
3. Relational model in 3NF.
4. EDM Relational structure
- 5.Volatile database with only latest transactions
6. DQ rules applied

Staging L5 – Relational ODS Permanent storage:

1. To store conformed and merged data in sequence from L4 using SCD rules.
- 2.Applies business rules on transactional data.
3. Relational model in 3NF and Type2 SCD historical data.
4. EDM Relational structure



Ent Data Architecture – Low level Details

Dimensional Data store – Star Schema Permanent storage:

1. To store aggregated, integrated ,consolidated ,summarised and historical data in denormalised star schema format .
2. To simple to understand and make it easy to find data that is required for end user. Optimized for large data sets.

Data Mart Layer:

1. To store subset or segment of enterprise datawarehouse relevant to a specific subject or domain
Eg: Procurement Data Mart,Sales DM etc..
- 2.To perform analysis faster on smaller subset of data.
3. Preferably star schema as a preferred data model to source PowerBI
- 4.Users are connected to Data Mart to manage data security and role-based access effectively.

Reference Database :

- 1.Reference DB contains Master data and unit of measure.
- 2.Master data in reference table gets validated and updated at staging L3 when data conformed and transformed entity by entity.

Data Quality Rule:

1. Reject the error records : This is when looking for accurate and complete data.
- 2.Accept the error records : This is when looking for completeness over accuracy.
3. Fix error records : To fix a specific issue by changing to correct value or format.
4. Assign default values : This is when looking for completeness by assigning a default value when the correct data cannot be found.

02 Enterprise Data Model



Components of Enterprise Data Model

The delivery, adoption and governance of **Data Services/Products** require maintained and accepted master data models and taxonomies

Subject Area Model

- The list of subject areas becomes one of the most significant enterprise taxonomies
- Subject areas typically share the same name as central business entities
- Important tool for the organization of governance, stewardship and data modelling.

Conceptual Data Model

- Defines business entities and their mutual relationships
- Business entities are named in business language and include a glossary containing business definitions and other Metadata
- Business entities will appear within the scope of several subject areas and should therefore have one primary subject area that 'owns' the master version of that entity

Logical Data Model

- Identifies the data needed about each instance of a business entity
- The essential attributes represent common data requirements and standardized definition for widely shared data attributes

EDM v1.3 5.pptx