Title: Common Sense Guide to Data Governance (Management)

Author: Rohit Singh

Table of Contents

1   Introduction

I have worked within the Financial Services industry for over 15 years primarily helping clients implement data solutions such as data warehouses and hubs. My recent roles have been focused on risk data systems to meet complex regulatory requirements such as the PRA's Firm Data Submission Framework (FDSF) and the BCBS 239 Principles of Risk Data Aggregation.

Senior managers primarily from Risk and Finance are increasingly asking me in private for a common sense explanation of Data Governance. They have had the presentations from large consultancies, data product vendors and the IT architects. The multiple terms and names have simply confused them and made them make investment decisions on faith that their IT and Change departments are correct rather than with confidence that they have understood what they are investing in.

This short paper is my Common-sense Explanation of why Data Governance is required and what the purpose of it is.

2   Why Do We Need Data Governance?

Over the past 15 years we have had a huge explosion in automation in the Financial Services Industry. This has in turn created its own challenges one of the challenges is the huge amount of data flowing across organisations. This data flow has been uncontrolled creating many challenges which are categorised below:

1. There is a lack of transparency to where data used for decision making comes from.

2. Often the data that is presented is not directly recognisable as to the context it applies to.

3. Data appears to be created spontaneously within an organisation and there is little control of the acquisition, production and consumption of this data resulting in serious risk to the organisation that cannot be identified.

4. Data flowing across the organisation is often ambiguous and open to interpretation as it traverses across the enterprise.

5. The quality of the data that is consumed by systems and finally used for decision making have a lot of reliability issues.

6. The data architecture of the business. Information systems have been rapidly developed and implemented across organisations without any form of overall control resulting in a

system and data landscape that cannot support the regulatory challenges that are increasingly being thrust on businesses.

3   Data Governance in a Nutshell

Data governance is the regulation and management of data flowing across the organisation so that the following objectives can be achieved:

1. Transparency; the data provenance (lineage) is evidenced by recording the data supply chain.

2. Recognisability; a data record when exchanged should have enterprise wide reference identifiers.

3. Orderly; data created spontaneously for departmental should be restricted for the intended use.

4. Consistent; any data exchanged between systems should be encoded in a common manner.

5. Quality that is fit for purpose.

6. Data architecture of the business should support the organisation to achieve 1 to 5.

3.1   Transparency; Document the Data Supply Chain

Financial organisations can be visualised as information processing factories where raw material is received at points in the manufacture of data it supplies to its stake holders such as clients, trading partners and regulators.

The transparency of this data supply chain across the business can provide revolutionary efficiency across financial organisations almost to the same level as that of process analysis introduced into manufacturing.

This supply chain can be documented by requiring every business unit to evidence Data Provenance. The data provenance principle states that for a data item to be provenance the following should be known (auditable)

- The source of data

- How the data were captured

- The meaning of the data that was first captured

- Where the data was stored

- The path of the data to the current location

- How the data were moved along that path

- How those data were altered along that path

This chain sometimes referred as data producer and data consumer links when joined provide data transparency.

3.2 Recognisability; Cross-Reference Data across the enterprise

Data such as Client or Product Identifiers are different as they flow across the enterprise this is a key challenge multiple links down a supply chain. A simple cross reference held centrally for reference (sometimes what is called master data) reduces this key challenge. This challenge is completely illuminated when all data producers are required to transmit data with identifiers that are standardised across the enterprise and even across the industry such as Legal Entity Identifiers.

3.3 Orderly; Establish Data Producer Consumer Contracts

Data created spontaneously for departmental should be restricted for the intended use. The formal way in which this can be established is for there to be a company policy that is auditable which established a Data Producer Consumer Contract. The objective of which is for the Data Producer to take on the responsibility of providing quality data and equally for the consumer to ensure that the data that is supplied meets the contracts.

3.4 Consistent; Data Is Exchanged in A Common Format

Data items are not exchanged across the enterprise in a consistent manner. To reduce this the enterprise could centrally hold a cross-reference table. However, this could be illuminated by requiring all data to be transmitted in a consistent manner across the enterprise.

3.5 Quality that is fit for purpose

Data Consumers should verify that the data supplied by data producers is fit for purpose across all seven data quality dimensions.

1. Accurate; Data accuracy can be proven by comparing or reconciling data against other sources to verify deviation. For example, a security price supplied from a back office can be compared with market data vendors such as Bloomberg to verify that the price is within permissible tolerance. Another way to ensure accuracy is to compare the difference between the last price to verify tolerance.

2. Integrity; A consuming system may be receiving the same data from different systems for example one feed may provide the portfolio names and another the portfolio positions. The consuming system should ensure that the portfolio names on the position feed match those transmitted in the portfolio names feed.

3. Completeness; A consuming system must verify that the data records it receives from source systems have all mandatory fields within the data contract populated.

4. Consistency; A consuming system must verify that the data supplied is consistent and encoded in accordance with the producer consumer data contract. This applies particularly to numeric values and how they are supplied here ambiguity can lead to serious financial loss.

5. Conformity; A consuming system must verify that the data supplied is encoded in accordance with the producer consumer data contract. Maintaining conformance to specific formats is important in data representation, presentation, aggregate reporting, search, and establishing key relationships.

6. Currency; A consuming system should verify that the data supplied is applicable to the time for which the data is required. For example, yesterday's price feed is not useful to value assets today.

7. Duplication: A consuming system should verify if there is multiple, unnecessary representations of the same data objects within the data feed. The inability to maintain a single representation for each entity across your systems poses numerous vulnerabilities and risks.

3.6 Data (IT) architecture to support Data Governance (Management)

There are IT tools available to support each of the previous sections and implementation of these can greatly aid Data Governance. However, there are dangers of investing in technology that does not provide an adequate return on investment. An evolutionary approach to IT is more likely to provide a return on investment the tools to be deployed in order are:

1. Metadata tool with a Graphical Interface capable of capturing data flows essential for representing the Data Supply Chain. Other facilities that the tool should have is to provide a glossary to record current data and ideal data standards.

2. The Data Architecture should be capable of managing business data flows both Batch and Message based in a holistic manner so that the data flow from producers to consumers can be traced in a timely manner. MQ message-based infrastructures have some key benefits for data governance when implemented correctly.

3. Data Quality tools that can be deployed on the data architecture to measure Data Quality across the seven data quality dimensions mentioned earlier. Data Quality Dashboards and data profiling tools placed strategically can assist greatly with data governance.

4 High level Data Governance Framework

4.1 Within the Governance framework of the organisation

1. Set up a data governance office to provide policies and procedures for documenting, monitoring and improving:

   a. The Data Supply Chain.

   b. Reference (Master) data standards across the supply chain.

   c. Data producer consumer contracts across the supply chain.

   d. Data exchange formats.

   e. Data quality measurement process at each data consumption point.

2. Disseminate data policies and procedures to business units and IT.

3. Implement the data governance policies and procedures through internal audit.

4.2 Establishing the Data Supply Chain

Prioritise and establish Critical Data Flows Dictionary

This exercise must be implemented to deliver value. As a result it is important to prioritise the data flows in terms of business criticality. These data flows called Critical Data Flows Transfer Critical Data Elements which are required for key decision processes. The output of this is a dictionary of Data Flows with the following assigned:

- Priority
- Data Flow Name
- Producer

- Consumer
- Business Owner
- Technical Owner

4.3 Establish Producer Consumer Contracts for Critical Data Flows

This exercise formalises the data flowing between the Producer and the Consumer. It defines:

1. Why the data is being transferred. What is the business purpose for the data flow and why is this flow required?

2. What data elements are being transferred.

3. Where the data elements are to be transferred.

4. When the data is required.

5. How the data is to be represented format and how it is transferred.

6. Data quality dimension required on the flow in terms of the 7 dimensions discussed earlier.

4.3.1 Document and Monitor Reference Data Standards on the Critical Data Flows

The dictionary of critical data flows can be enhanced with the reference or master data used to identify the data records being transferred such as portfolio code or security identifier. These items are defined in detail at a later stage.