VANQUIS Banking Group

Data Governance Frameworks

13th March 2024

Garry Manser

Group Head of Data Governance

DATA GOVERNANCE AND MASTER DATA MANAGEMENT CONFERENCE EUROPE

11 - 14 March 2024 |London, UK

Please score and comment on this session and speaker in the event mobile app

IRMUK

And who am I?

I am currently the Head of Data Governance at Vanquis Banking Group and am accountable for the design, development, and implementation of the Group's Data Governance Framework, specifically enabling Data Owners and Data Stewards to understand their data more, whilst the Group embarks on a significant Data Transformation programme.

I have been working in Data Governance for too many years, both in Consultancy for EY and PWC, and in several financial institutions such as Barclays and Visa. Through that I have had a number of adventures and I could write a book on the bad mistakes made and lessons learned.

My Journey in Data Governance

VANQUIS Banking Group

Who are we?

We're Vanquis Banking Group plc – A leading specialist bank, focused on the near prime and mid cost credit markets.

The Group has been providing financial inclusion, which supports social mobility, to consumers whose needs are not well met by traditional lenders for over 140 years. We lend responsibly and provide tailored products and service propositions to 1.7 million customers throughout the UK.

We're a FTSE 250 company, proud to support the 1 in 5 people in the UK who can't get access to credit products through mainstream banks and building societies. We want to provide a helping hand when others don't. That's why our aim is to continue to develop better banking products for the 10-12 million people in the UK who deserve good quality products that they can't get anywhere else.

Our purpose: To help put people on a path to a better everyday life.

We offer credit cards and loans as Vanquis and Vehicle Finance as Moneybarn. All are delivered online and are built with customer flexibility in mind.

What will we cover today?

- Why a Data Governance Framework ?
- What makes up a good Framework?
- Deep dive into each of the sections.

- With a few stories on the way.

VANQUIS Banking Group

Why a Data Governance Framework?

How do we ensure our Data is fit for purpose?

Where many of us started?

Where many of us went wrong.

The Response.

The Response becomes.

VANQUIS Banking Group

The Data Governance Framework

How do we ensure our Data is fit for purpose?

Data Governance Framework

The framework delivers the scope of the Data Governance programme.

It allows Teams to understand the full picture without the “must” statement Day 1

The Framework will also support the following areas.

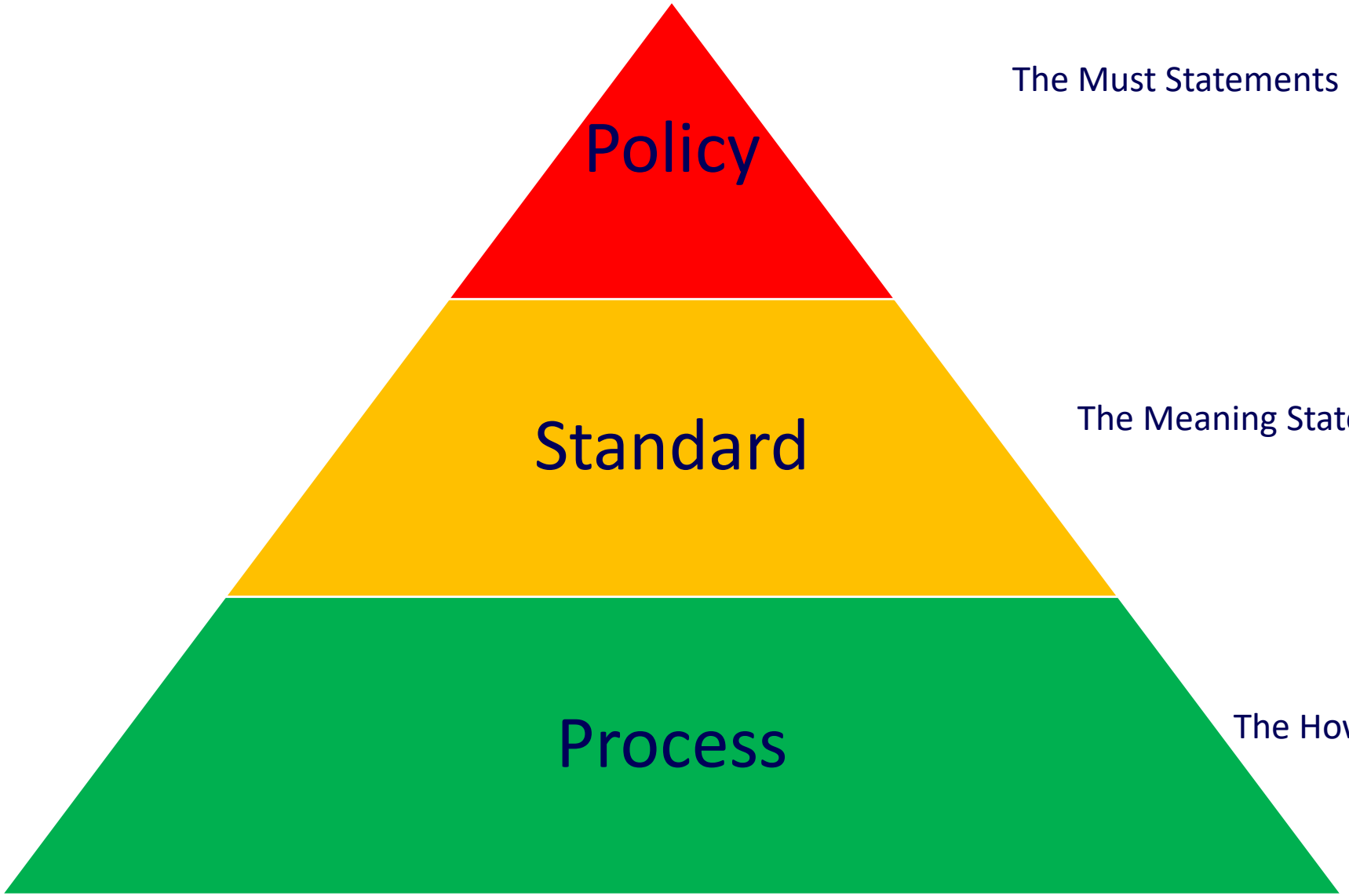
- Data Protection and Privacy.
- Information Security.

The Framework introduces the concept of Governance by Design.

Where to start.

- We all tend to start with the Policy – don't start with the Framework and then morph that to the "Must" Statements
- Start Simple and don't over commit – these all take time
- Data Quality, Master and Reference Data and Records Retention are good standards to start with, the rest can be covered in the policy
- As you formulate the framework, ensure that you plan when these are to be delivered.

Policy , Standard, Process



The Must Statements

The Meaning Statements

The How Statements

Why Data Quality?

- Ensure you are able to articulate the purpose of Data Quality controls, focus on benefits.

Don’t do this!

Me!

The Door

What is Data Quality Management?

- Data Quality Management is a perception of an assessment of data's fitness to serve its purpose in a given context.
- The objective is to understand the quality of data, and where appropriate fix any data that is incorrect, inaccurate, incomplete, formatted incorrectly, duplicated, out of date or irrelevant to the data set.
- This is accomplished by replacing, modifying, correcting, or deleting any data that falls into these categories.
- However, it is important that the root cause of data quality issues is understood prior to any remediation actions.


No entry for heavy
goods vehicles.
Residential site only
Nid wyf yn y swyddfa
ar hyn o bryd. Anfonwch
unrhyw waith i'w gyfleithu.

What are Data Quality Dimensions?

Data Quality Management is delivered through a series of dimensions. Not all dimensions may be applicable to a data object.

What are Data Quality Dimensions? The ones we use.

Data must be accurate.
Is the data a true reflection of what is required? Accuracy is defined as “the extent to which a data attribute, object or data set correctly represents the real-world property or item it is meant to represent.”

Data must be consistent.
Is the data attribute or object consistent across multiple data sets? Is there a golden source / record defined for authentication? Consistency is defined as “the extent to which a single attribute has the same value across all the data sets in which it appears”.

Data must be valid.
Does the data meet particular rules or constraints as defined? Validity is defined as “the extent to which a data attribute’s value conforms to a defined set of business rules.”

Data must be complete.
Are there missing elements in the data set? Completeness is defined as “the extent to which all data attributes are defined a value in the data set.”

What are Data Quality Dimensions? The ones we use.

Data must be unique.
There should be no duplicate data objects, attributes or records in a data set. Uniqueness is defined as “the extent to which no specific record or entity appears more than once in a data set.”

Data must be timely.
Rules to declare that a record is obsolete or inactive are defined per object, for example “no activity in the past n years”. Timeliness is defined as “the extent to which data is sufficiently up to date for the use at hand.”

Volumetric controls should be in place.
Rules ensure that data received is in line with expectation, for example transaction numbers based on calendar and expectation. Volumetricity is defined as “The extent to which data is consistent across a time series or are there expectations that data will rise and fall subject to external factors, for example the day of the week.”

Reporting Data Quality?

- Consider how you want to report your data.
- Ensure Data Quality thresholds are agreed prior to starting.

Which would you use?

Your data is 95% Complete.

Your data is 5% Defective.

Which do I use?

~~Your data is 95% Complete.~~

Your data is 5% Defective.

How important is Data Cleansing?

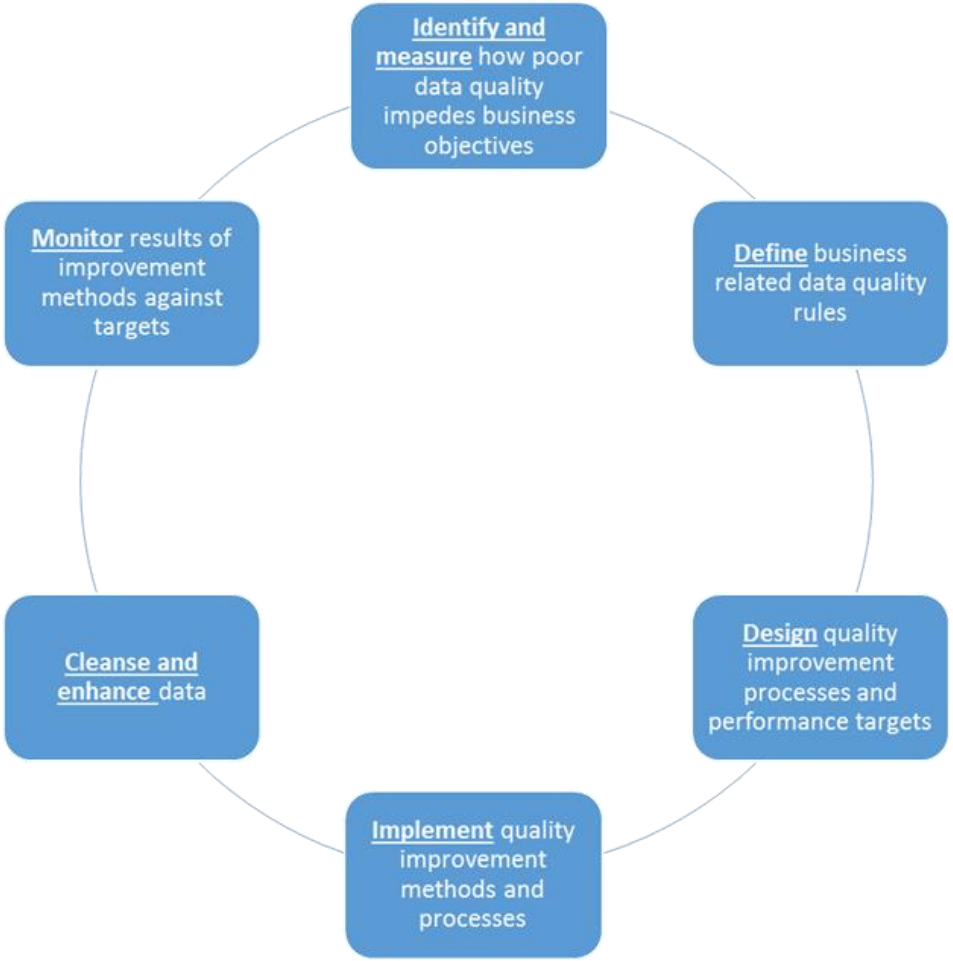
- The objective is to understand the quality of data, and where appropriate fix any data that is incorrect, inaccurate, incomplete, formatted incorrectly, duplicated, out of date or irrelevant to the data set.
- This is accomplished by replacing, modifying, correcting, or deleting any data that falls into these categories.
- But there are pitfalls in doing this

How important is Data Cleansing?

- The objective is to understand the quality of data, and where appropriate fix any data that is incorrect, inaccurate, incomplete, formatted incorrectly, duplicated, out of date or irrelevant to the data set.
- This is accomplished by replacing, modifying, correcting, or deleting any data that falls into these categories.
- But there are pitfalls in doing this

- Always look to ensure that you have identified the issue and remediated, prior to looking to cleanse the data.

Data Quality Management Circle



Comfort Break

Master Data

- The Data you require to run your business
- May be Customer Data, Supplier Data, Employee data
- Principles – Capture Once – Store Once – Use Many times

- Likely to be all over your organisation – can you identify the Golden Source ?
- Back to Data Quality - Consistency

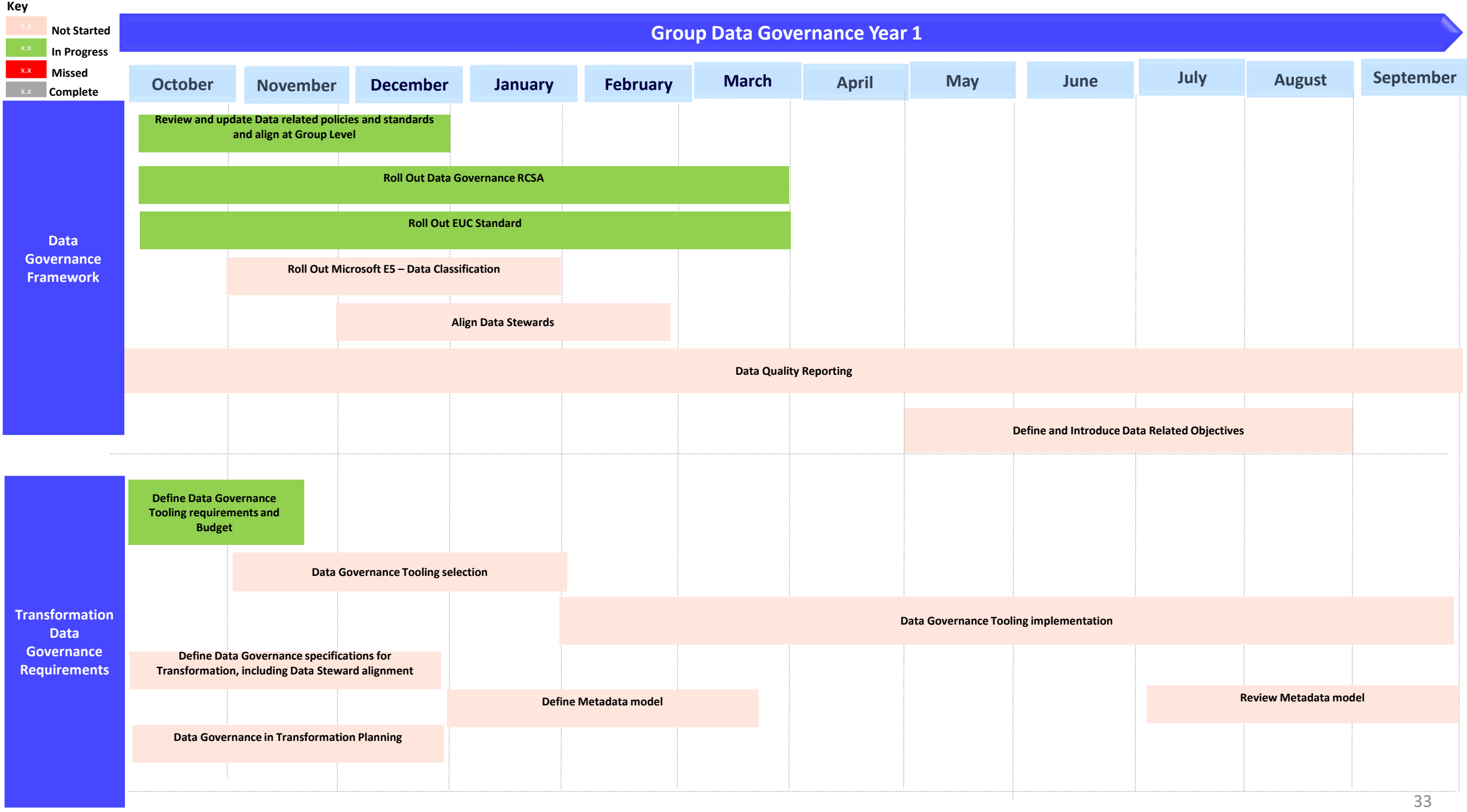
Reference Data

- The data you use regularly, likely to be lists
- May be Country Codes, Currency Codes
- Principles – Capture Once – Store Once – Use Many times

- Likely to be all over your organisation – can you identify the Golden Source ?
- Can you update sources?
- Back to Data Quality - Consistency

Consider your Strategy – and Other Company Strategies

- How do you intend to deliver?
- What other initiatives can you assist?


Key
x.x Not Started
x.x In Progress
x.x Missed
x.x Complete
Group Data Governance Year 1
October
November
December
January
February
March
April
May
June
July
August
September
Data Governance Framework
Review and update Data related policies and standards and align at Group Level
Roll Out Data Governance RCSA
Roll Out EUC Standard
Roll Out Microsoft E5 – Data Classification
Align Data Stewards
Data Quality Reporting
Define and Introduce Data Related Objectives
Transformation Data Governance Requirements
Define Data Governance Tooling requirements and Budget
Data Governance Tooling selection
Data Governance Tooling implementation
Define Data Governance specifications for Transformation, including Data Steward alignment
Define Metadata model
Review Metadata model
Data Governance in Transformation Planning

It’s all about the people

Data Owner Accountabilities

- Takes overall accountability for their function's data sources. This includes ensuring their data's adherence to the Group's data governance standards and principles.
- Accountable for directing access management rules specific for their data.
- Accountable for ensuring that their data is secured with appropriate controls, proportional to its sensitivity.
- Accountable for the correct completion of data creation, data updates and data deletion in addition to championing data standards.
- Overall accountability for continually ensuring that their data is fit for purpose, complete, accurate, reliable and available on a timely basis.
- Responds to monitoring reports across data governance by taking appropriate decisions to address and resolve issues as required and on a timely basis.
- Reviews the regular reporting packs and provides input in advance of the relevant data governance meetings as required.
- Ensures that the ROPA (Record of Processing Activities) is updated and reviewed on a periodic basis.

Data Steward Responsibilities

- Responsible for prioritising Data Governance efforts across their respective function, ensuring colleagues are aware of their own responsibilities for data.
- Attends or sends a delegate to the Data Steward Forums, responsible for sharing changes and actions with colleagues where appropriate.
- Responsible for ensuring the Business Glossary, Data Dictionary and Data Catalogue are in place, up to date and complete for all data in their respective function
- Responsible for ensuring Records Retention schedules are adhered to and data moved/deleted as appropriate
- Day-to-day ownership, maintenance, and management of the respective function's Data Governance RCSA ensuring that process level risk exposures remain appropriately rated both on a Gross and Net basis, and that the function's data governance controls remain effective.
- Supports the definition of the scope for new data initiatives and data definitions.
- Supports the definition of Data Quality rulesets across all data dimensions and implements dashboard reporting of same.
- Drives the investigation and resolution process for data quality issues, including process changes where appropriate.
- Champions and supports Data Privacy activities across their respective Business Unit, including but not limited to completing DPIAs (and identifying where required) and identifying and reporting personal data breaches. Acting as the risk delegate where allocated in the RCSA for DP risks and controls.
- Responsible for authorisation of data access requests for data they are responsible for, ensuring that access is only provided where necessary.

Data Consumers and Technical Custodians

A Data Consumer is an individual who receives and uses data to conduct VBG business activities. This applies to all data consumption activities, including system feeds, 3rd party data, and those activities whereby a colleague directly accesses data from a system or via an interface.

A Technical Custodian is accountable for the technology that creates, maintains, and distributes data throughout Group.

- Documenting the systems in which data is stored.
- Ensuring data access requirements are delivered.
- Ensuring all changes to systems are audited and communicated to Data Stewards in advance to analyse impacts.
- Provide effective, audited transfers of data between systems where required.

What makes a good data Steward

The Data Steward
This role is the combination of new and existing responsibilities, developed in order to improve the management of data across the bank. It is a part time incremental responsibility and is likely to be aimed at Senior Manager level with support from functional colleagues.

The right person will be identified as having:

- Good understanding of their local data, the requirements from their team, where data comes from and how to manage it well, to ensure quality.
- Ability to explain data issues in simple layman's terms.
- Enthusiasm and vested interest in improving the business.
- Well known in their team, often the go-to person for data issues already (such as reporting or quality concerns).

Their primary responsibilities include:

- Being a point of contact for queries about policy and good practice for their teams.
- Maintaining an accurate and complete record of Data Assets, Sources, and Ownership; being consulted on any additions or changes to assets they oversee.
- Approving access requests and reviewing access regularly to ensure those who no longer need it have their privileges revoked.
- Periodically review local data against retention policy, ensuring the approach to archiving, retaining, and deleting data is followed.
- Representing their team for data queries, including attending forums, escalating issues, and advising on changes to future policies, practices and training.

The Data Steward will not be expected to be an expert on latest data regulation, and will receive guidance from the Data Protection Officer on related legislation and appropriate data practices

Deborah is a senior manager working in the Marketing team, she manages the CRM including monthly reporting on the success of email and social media campaigns. Over time, she has become responsible for reporting quality issues to the input staff, revising training to ensure consistency. She is usually also the go-to person to request access to systems and data.

Harry is a manager working in the HR department and is responsible for HR Operations. He manages all data required for a colleague's lifecycle from hire to retire and is often called upon to provide reports on resource headcount, sick absence, and annual leave. He is additionally responsible for joiners and leavers and often provides updates to other functions.

Cam is a team leader in Client Services and is responsible for the maintenance of client accounts and records. They deal with customer account queries, as well as deal with account opening and maintenance. As expected, the team is on the front line for managing client account data, and Cam ensures data quality is accurate and legitimately held across all systems.

Though we need tools too

What is a Data Dictionary?

As the name suggests, Data Dictionaries provide information about your data. Descriptions can include data attributes, fields, or other properties such as data type, length, valid values, default values, relations with other data fields, business definition, transformation rules, business rules, constraints etc.—anything you need to define each physical data element inside operational data sources and data warehouses. This is also relevant for logical BI data objects, and it should have a business flavour to it, not just technical.

A Data Dictionary should be a one-stop-shop for IT system analysts, designers and developers to understand everything about their metadata. They are used to help translate data level business requirements into technical requirements and should ideally be able to provide this information in an easily understood, structured and organised way. IT teams should be able to tell within a few seconds exactly which inputs should be included to meet project goals, from attribute type to field requirements to default values.

What is a Data Dictionary?

Data Dictionaries are often presented in spreadsheet format with rows and columns defining each attribute or metadata category that needs to be addressed in a system. They are sometimes something someone enters on their own and has to comment on and refresh manually. Data Dictionaries look over the system catalogue of a database and pull specific objects into the database. For a column this may include:

Column Name

Column Location

Column Datatype

Descriptive information that a user has entered available in the System Catalogue

Information within a Data Dictionary mainly helps BI developers. The Data Dictionary is essentially an inventory that shows which type of tables and columns exist.

So what then is a Business Glossary?

Business Glossaries help define terminology across business units. They offer clear definitions across the entire enterprise with the goal of keeping terms consistent and helping everyone stay on the same page.

A quality Business Glossary is an important part of collaboration, particularly in larger businesses that span numerous departments. You'd be surprised at how differently each different business unit defines data elements relevant to their own operations, even in related departments (such as sales and marketing). As self-service users define the logical meaning of data elements and can create their own calculated columns, there is a lot of room for inconsistency.

And a Data Catalogue?

Much like the BI team's role in creating one source of the truth in the data, Data Catalogues provide one source of the truth about the data. While Business Glossaries help define terminology across business units and Data Dictionaries provide technical information about physical data assets, Data Catalogues are a one-stop shop for anyone shopping for data they would like to use, manage or understand.

The data catalogue ties the business terms to the physical data assets and includes capabilities intended to make organisational data easy to locate, understand and use. Having good business definitions is great but if they are not related to their underlying data their value is greatly limited, when any user wants to locate data associated with a business term they will need to start hunting for it with the BI team. Good technical documentation is great for the technical team, but it will only take BI so far, the BI team will most often still need to spend a majority of its time mitigating between the data users and the data. True independence in using the data to its fullest can only be achieved by fully democratising the data to all data users through documentation that bridges the gaps between technical, physical and semantic data assets and their related business terms and aspects.

So its all about the Data Catalogue?

Data driven organisations have realised the necessity of empowering data users, doing this starts with organisational wide accessible documentation and collaboration around the data, these organisations recognise that an excellent Data Catalogue is the most efficient way to unleash the power and hidden value of all the data they work so hard to create and maintain.

With a quality Data Catalogue, any data consumer should be able to easily locate any data asset, understand what the data means, how the data can be used, what are its limitations, who is accountable and responsible for the data, what else the data is associated with and last but not least, enable in context collaboration around the data.

A (stolen) example

	Content	Example
Business Glossary	• Terms relevant to the business with a unique definition for each • Each term's involvement in KPIs and other metrics • Cross-referencing of terms and relationships	A Social security number (SSN) is a unique number assigned by the US Government to US Citizens for the explicit purpose of identifying individuals within the Social Security system. The use of SSNs has grown within the US Government and is now used as a unique identifier for income tax and other purposes. Use of SSN as a unique identifier has grown beyond the US government and is typically used in the private sector to uniquely identify individuals for many purposes such as credit reporting, health care, etc..
Data Dictionary	• Schema, table, column, data type, triggers, constraints, etc. • Allowable values, ranges, related reference data, etc. • A mapping of each data element to the relevant business glossary terms • Each data element's involvement in KPIs and other metrics	SSN itself is a nine character string and is typically displayed with hyphens between the 3rd and 4th characters and 5th and 6th characters as follows: xxx-xx-xxxx. Each character other than the hyphens are numeric: 0-9. SSN is stored as a character string rather than a numeric data type due to the common occurrence of leading zeroes in many SSNs. This leading zero would be automatically dropped if stored in a numeric field thus affecting the integrity of the datastore. Databases might store SSN in a string of 9 character length or with the hyphens in a string of 11 character length. SSN is a unique identifier and can be used to join across tables in many systems. SSN is considered PII and/or PHI depending on context and is to be handled with appropriate levels of security measures.
Data Catalog	• Directory including the location(s) where each specific data element or grouping of elements exists • Explanation of any redundancy • Cross references the business Glossary to the various Data Dictionaries	SSN is stored in the <table-name> in the <data-source> and also in the <table-name> in the <data-source>. Both systems store SSN in 9 character varchar and mask the first 5 numeric characters showing only the last 4 characters as needed in the following typical format: 'xxx-xx-1234'. Full unmasked SSN is viewable by resource with a business need to see this information. Cross reference to <business glossary definition>

Who can see what

Why is Data Access Important?

Access controls limit access to information and information processing systems. When implemented effectively, they mitigate the risk of information being accessed without the appropriate authorisation, unlawfully and the risk of a data breach.

So what goes wrong?

Historically organisations have taken a lax approach to Access Management, looking for easy options.

What I have seen many times.

When someone new started, an organisation would look at what data access that person required, and would usually look at who had been in the team the longest and copy their access profile. But that person had usually been in the Organisation a long time, would have worked in other departments and would have access to data they no longer required, but the new starter also received all that access.

So what is Role Based Access?

Role-based access control (RBAC) refers to the idea of assigning permissions to users based on their role within an organisation. It offers a simple, manageable approach to access management that is less prone to error than assigning permissions to users individually.

When using RBAC, you identify the needs of your users and group them into roles based on common responsibilities. You then assign one or more roles to each user and one or more permissions to each role. The user-role and role-permissions relationships make it simple to perform user assignments since users no longer need to be managed individually, but instead have privileges that conform to the permissions assigned to their role(s).

For example, if you were using RBAC to control access for an HR application, you could give HR managers a role that allows them to update employee details, while other employees would be able to view only their own details.

What are the benefits of Role Based Access?

With RBAC, access management is easier as long as you adhere strictly to the role requirements. RBAC helps you:

- create systematic, repeatable assignment of permissions
- easily audit user privileges and correct identified issues
- quickly add and change roles, as well as implement them across APIs
- cut down on the potential for error when assigning user permissions
- integrate third-party users by giving them pre-defined roles
- more effectively comply with regulatory and statutory requirements for confidentiality and privacy

And the Data Governance aspect?

Data Owners are accountable for controlling access to the Data they own. Access granted should be relevant to the role the individual Data Consumer is required to undertake, general access to data should not be permitted. Role Based Access Controls should be defined across the Group, with specific consideration given to the joiners, movers, and leavers processes. Access controls should also undergo periodic reviews (maximum annually) to ensure that Group data is held securely.

Additional consideration should be given to access granted to PII data, in line with Regulatory requirements.

These controls should be reviewed on an annual basis and should be linked to JML (joiner, mover, leaver) processes to ensure relevant updates are made as required.

How much data to keep?

Records and Data Management is the systematic control of data assets throughout their lifecycle in order to meet the operational needs of the business as well as statutory, regulatory and fiscal requirements. Effective data management allows fast, reliable and accurate access to data, ensuring that it is deleted at the appropriate time. It is not a reason to keep everything forever, or to apply a standard 7-year retention to everything held.

How much data to keep?

Consider how you educate the Business.

- Team Town Halls
- Articles
- Specific Training

Thanks for listening!

VANQUIS Banking Group

DATA GOVERNANCE AND MASTER DATA MANAGEMENT CONFERENCE EUROPE

11 - 14 March 2024 |London, UK

Please score and comment on this session and speaker in the event mobile app

IRMUK