Data Stewardship

- Companies are getting serous about managing their data, including improving data quality, understanding the meaning of the data, leveraging the data for competitive advantage, and treating data as the enterprise asset it should be.
- But doing a proper job of managing data requires accountability – that is, business functions must take responsibility for the data they own and use. The formal recognition of the need to have a structure, organization, and resources in place to manage data and the actual implementation of that need – has come to be known as data governance.

Problem statement

- data doesn't' explain itself
- data is shared and used by many, for many different purposes
- many processes that use data depend on people upstream of the process to "get it right"
- the software development life cycles require many handoffs between requirements, analysis, design, construction and data use
- technical people tasks with data implementation are not familiar with the data's meaning or how it is used

- All of these factors lead to poor understanding of the data and a perception of poor quality, with little ability to know the difference.
- Some of the failures include:
    - **Data definitions**
        - Often written in haste by project staff and are not rationalized across the enterprise
    - **Data quality**
        - Rules are often not defined and the quality itself is rarely measured
    - **Documentation**
        - Documents containing metadata are rarely officially published
    - **Creation and usage business rules**
        - A lack of understanding about the conditions under which an entity can or should be created, as well as how data should be used
- Like other assets, the data needs to be inventoried, owned, used wisely and understood
- Establishing ownership requires understanding how the data is collected and who uses it, then determining who can best be responsible for the content and quality of the data elements.

What is data governance?

- All about making sure that people are properly organized and do the right things to make their data understood, trusted, of high quality, and ultimately suitable and usable for the enterprise's purposes

What is data stewardship?

- An approach to data governance that formalizes accountability for managing information resources on behalf of others and for the best interests of the organization.

Overall goals of data stewardship

- Processes and procedures are written, approved and in use to:
    - Identify key business data elements
    - Collect, review, and approve business metadata for the key business data elements
    - Log, analyze, prioritize, and remediate data and data quality issues
    - Support projects
    - Manage domain data
    - Replace data stewards as necessary
    - Review and analyze data quality improvement opportunities
    - Document stewardship work done and methods of publication

Moving data to a governed state

- Fully governed data means that you know all of the following at the data element level:
    - The standardized business name of the data element
    - The standardized business definition of the data element
    - The calculation or derivation rule
    - Physical location of the business data element
    - The data quality rules, in context
    - Rules for creating the data element
    - The data governor/owner and business data steward for the data element

Three P's: Policies, Processes, and Procedures

- To have an effective data governance and data stewardship, it is necessary to have three Ps
- Data stewards don't write the policies, but they must be aware of what policies have been written
- The data stewards create the metadata and make recommendations for the data governors to act on

Executive steering committee

- Drives cultural changes needed to treat data as an asset and manage it effectively across business-area boundaries
- Makes necessary changes to the organization and tools as required for effective data governance
- Creates and promotes the vision for the data governance program
- Authorizes the data governance board budget
- Balances business priorities with operational needs across the enterprise
- Approves data governance policies
- Reviews, evaluates and reports to executive sponsorship on the data governance performance and effectiveness
- Provides advice, direction, counsel, and feedback to the data governors (members of the data governance board)
- Ensures decisions regarding the data support the strategic direction of the organization
- Ensures active participation by the business and IT
- Represents their business function's direction and views in regard to the adoption and deployment of enterprise policies and practices
- Appoints a data governor from their business functions
- Resolves issues escalated by the data governance board

Data governance board

- Primary decision makers, acting on recommendations from the data stewardship council
  - Has funding authority to spend budget money on data management improvements
  - Prioritizes decisions regarding data to address the most relevant needs of the organization
  - Reviews, evaluates, and reports to the executive steering committee on data governance performance and effectiveness
  - Ensures that annual performance measures align with data governance performance and business objectives
  - Reviews and approves data governance policies and goals
  - Ultimately is accountable for business data use, data quality, and prioritization of issues
  - Make strategic and tactical decisions
  - Reviews and where appropriate, approves the recommendations made by members of the data stewardship council

- Assigns the business data stewards to the data stewardship council
- Represents all data stakeholders in the data governance process. Ensures appropriate representation and participation in data governance across the enterprises
- Owns and governs the business processes that produce data
- Identifies and provides data requirements that meet both enterprise business objectives and the objectives of their business function
- Defines data strategy based on business strategy and requirements
- Communicates concerns and issues about data to the data stewardship function

IT support through technical data stewards

- The primary role of the IT technical data stewards is to provide technical expertise in support of the data governance efforts with respect to systems and application impact analysis for proposed changes and data quality issues

Data governance program office

- Supports, documents and publishes the activities of the executive steering committee, data governance board and data stewardship council
- Defines and documents best practices in data governance

Choosing the right business data steward

- Answer the question once
- Fewer arguments
- Repeatable process of changes
- Additional rewards and recognition

- Master data (data for which multiple instances of the data exist across the enterprise and for which it is advantageous to have a single reconciled or “golden” copy) is often described as a data domain.
- You have to be able to state who the business data steward is down to the data element level.
- Technical data stewards provide support and are associated with specific systems, applications, data stores and technical processes, such as identify resolution (for master data management, MDM), data quality rules enforcement and ETL jobs.
- Project data stewards can be analysts already assigned to a project, a member of the data governance program office, or even a contractor
- Project data steward is not responsible for the decisions about the data. This distinction is important, because once a project data steward gains some experience and gets to know the data, there can be a tendency to start making decisions on his or

her own. Any decision made by the project data steward does not carry the authority of the business data steward or the data governance program.

- Project data steward is responsible for:
    - Recognizing and recording issues and questions that require the input of the business data steward
    - Coordinating the list of issues and questions with other project data stewards, because multiple projects may well come up with the same issues and questions
    - Presenting the list to the appropriate business data stewards and recording decisions and input
    - Providing the results as input back to the project

Responsibilities of the data stewardship council

- Focus on ways to improve how an organization obtains, manages, leverages and gets value out of its data
- Be the advisory body for enterprise level data standards, guiding principles and policies
- Mediate or arbitrate the resolution of issues
- Communicate decision of the data stewardship council and data governance board
- Ensure alignment of the data governance effort to the business
- Participate in and contribute to data governance processes
- Communicate the data governance vision and objectives across the organization
- Communicate the rules for using data
- Review and evaluate data governance performance and effectiveness
- Provide input into data governance goals and scorecard development
- Collaborate with other interested parties in the management of definitions, policies, procedures and data issues
- Enforces use of agreed-on business terminology

Enterprise data steward

- Report up to the data governance manager
- Lead the community of data stewards
- Liaise with data governors/business leads or their appointees as well as IT project leads to implement and maintain data governance
- Work with the data governance manager to develop the data governance vision and framework, short term and long term
- Identify and initiate projects to implement the vision

- Ensure all data governance work efforts are in line with overall business objectives and the data governance vision
- Manage the data domain stewards
- Define prioritization criteria
- Provide direction to business and IT teams
- Lead implementation of data stewardship organization
- Design the processes and procedures for data stewardship
- Build and drive the agenda for data stewardship council meetings
- Maintain a repository of information and decisions
- Improve overall enterprise data quality and reliability through process improvement
- Review and manage issues, and meet with the business to understand user needs and technical feasibility
- Work with data governors and data stewards to facilitate the issue resolution process
- Provide counsel to projects to ensure the project is in line with the vision of the data governance program
- Define, implement, and manage data governance metrics
- Track, monitor, and publish data governance scorecards

Business data stewards

- Member of the data stewardship council
- Aligned to a business function
- Responsible for data governance execution in their functional area
- Identify key business terms, data quality requirements, and data quality metrics
- Provide input for data governance metrics
- Represent interests of data governors
- Work with data governors to ensure business has a practical understanding of the data
- Participate in process and standards definition
- Ensure that data decisions are communicated and business users understand impacts of the decisions to their lines of business
- Provide business requirements on behalf of aligned function

Data life-cycle management

- Facilitate the data governors through the change control process
- Coordinate business requirements and requests specific to stewarded data area
- Define the business data definitions and appropriate data usage
- Own data metrics for compliance with data governance policies and standards

- Participate in conflict resolution
- Assess enterprise impacts related to data changes
- Organize and participate in data stewardship committees
- Work on behalf of data governors to ensure consistency of data usage and share best practices
- Work with the business stakeholders to define the appropriate capture, suage, derivation, and data quality business rules for all governed data elements within their data areas
- Assist in the identification, definition, and population of the correct and required metadata

Domain data stewards

- Aligned to a specific domain of shared data
- Represent the interests of the enterprise-wide use of the data
- Make recommendations regarding data that is used across multiple business functions
- Build a consensus around domain data usage among the data users across the enterprise
- Work with business data stewards, project data stewards, and data governors to ensure business and projects have a practical understanding of the data
- Participate in processes and standards definitions
- Work with the business to define acceptable levels of data quality
- Gather and report data quality metrics, and define improvement opportunities
- Defining a valid list of values for data elements used across the enterprise
- Maintain the names and descriptions of the data elements being used in a project
- Work with business data stewards to determine the appropriate business data steward to take responsibility for the project data element
- Review the name and description with the appropriate business data steward and get a business definition
- Collect and document business derivations and calculations
- Review the derivations and calculations proposed with the appropriate business data steward
- Deliver business data steward decisions to the project for incorporation in the project plan
- Collect and document data quality rules and data quality issues from the project
- Evaluate the impact of the data quality issues on the project data usage and consult with the data governors or data stewards where appropriate

- Consult with the business data steward and project manager to determine if data should be profiled based on data quality rules and expectations collected on a project
- Assist the data profiling efforts by performing data profiling tasks
- Inform and consult with other data stewards, data governors, and the enterprise data steward about definitions and data quality rules and issues that result from a project
- Work with project managers and project members throughout the course of a project in a collaborative manner while ensuring the data governance-related concerns are addressed for each project
- Where possible, align with projects that utilize a project data steward's previous experience and expertise

Technical data stewards

- Provide the technical expertise around source systems, ETL processes, data stores, data warehouses and business intelligence tools
- Explain how a system or process works
- Check code, copylibs, internal database structures and other programming constructs in search of how the information is structured, how the data moves, and how the data is transformed within a system or between systems

Operational data stewards

- Ensure adherence to data creation and update policies and procedures while creating new values or modifying existing ones
- Assist business data stewards in identification and collection of data metrics
- Assist remediation project team with changes to the data, application processes and procedures
- Assist business data stewards in performing data analysis to research issues and change requests
- Identify and communicate opportunities for data quality improvements

- Business data stewards bear primary responsibility for the data owned by their business function, are supported with some of the hands-on work by operational data stewards, and depend on technical data stewards for technical information. In addition, data stewards have responsibilities as a team (the data stewardship council)

Championing and communicating data stewardship

- If the data analysts are aware of data stewardship and how it operates, they can:
    - Bring up matters or data issues to the appropriate people

- Offer solutions and known workarounds for data issues
- Make available data dictionaries, lists of valid values, queries, data quality specifications, and other artifacts that have been useful to them
- Advise people to contact the data stewardship function when the analyst believes that a data stewardship procedure is not being followed

Gaining support from above and below

- Data stewardship often involves cultural changes, and only the executives can make those happen. Finally, if the rewards system is going to be adjusted to encourage active participation in data stewardship, high level management support is needed for that too

Organization structure

- The first step in sorting out the organization's decision-making levels is to understand how the organization is structured
- Focus on the business functions. Focusing on business functions means that restructuring is largely irrelevant.
- For an example, a large bank grew by acquiring other banks, and keeping systems and business functions that were superior to the existing systems and functions. This lead to the odd situation where first mortgages were issued from one subsidiary, second mortgages were issued from another, and personal credit was issued by a third.
- As you can probably imagine, the vast majority of data used by the three systems was the same, yet issuing a combined first and second mortgage was next to impossible.
- Each business function felt that it owned the data it used, and the only way to resolve this was by joint ownership, with the three data stewards working closely together to reach a consensus on meaning and business rules.
- A side benefit of this arrangement was that the three stewards were able to determine in which systems the data was or better quality and match customer/account holders across the systems.
- This was possible because in a banking environment, the SSN was a valid identifier for the account holders and was always collected when the account was opened

Organizing data stewards

- Business functions can be broken up into two subtypes: those that own data and those that don't
- Business data elements can be broken up into two subtypes: those that can be stewarded by a single business function and those that must be jointly owned by multiple business functions

- A data-owning business function is represented by a business data steward, who is assigned by a representative from the data governance board.
- Projects have or define business data elements, which are shepherded along by the project data steward
    - Determine ownership for new data elements
    - Coordinate the work between the business data stewards and project data stewards
    - Determine interested parties for shared data domains
    - Prioritize new data elements to be brought under data governance

Figuring out what you've got: Data

- What data you have
- Where the data comes from
- What parts of the organization are responsible for source data capture, ETL, and reporting?

Figuring out what you've got: The Metadata

- Metadata is key to a data stewardship effort. Much of what you need to discover and document about your data is its metadata
    - Finding definitions
    - Finding derivations
    - Where is the metadata being kept?
    - Is anyone managing and validating the metadata and by what means?
    - How is IT tracking its metadata?
    - Are there any data dictionaries for legacy systems lying around?
    - Does the current project methodology support capture and validation of metadata?

Figuring out what you've got: Data Quality

- What data quality issues have been raised and documented?
- Are there any projects "in flight" to fix data quality issues?
- Does the current project development methodology support capture of data quality rules?
- Is anyone doing data profiling?

Training for data quality improvement

- Framework for data quality:
    - How the organization defines quality data?

    - What data quality rules are, and how to define them
    - Detection and documentation of data quality issues

Practical data stewardship

- The key is to focus on just the practical and fundamental aspects of data stewardship: determining the key business data elements and assigning stewardship for them, creating quality metadata, creating and following a set of repeatable processes, and putting procedures in place to streamline the logistics of working together.
    - Financial reporting data
    - Compliance and regulatory data elements
    - Data elements introduced by company executives
    - Data used by high-profile projects
    - The business data stewards decide

Characteristics of a good business definition

- The definition of the term, using business language
- What purpose the term serves to the business – how the business uses the information represented by the data element
- Must be specific enough to tell the term apart from similar data elements
- Should link to already-defined terms that are used in the definition
- Should either state the creation business s rules or link to them

Setting up repeatable processes

- Bringing new data elements under governance
- Managing the business glossary
- Evaluating and finding resolutions for data quality issues
- Resolving a data governance request or issue
- Managing policies, procedures and metrics
- Coordinating the work of multiple project data stewards
- Managing the issue log

Aligning reference data values across systems

- Data elements with the same name that mean something entirely different
- Data elements with different names that have the same meaning
- Data elements that mix several different elements together

Operational metrics

- The level of participation of the business

  - The level of importance given to the data stewardship effort
  - Numerical measures for results achieved
  - How often and how effectively the data stewardship deliverables are used
- Metrics are important to show the value that data stewardship adds to the enterprise, and to therefore justify the resources and efforts put into the data stewardship effort. Business value metrics show how the data stewardship program contributes value to the company in terms of increased profits, reduced costs, shortened time to market, fewer compliance and regulatory issues and other measures that translate into conducting business better and more efficiently.
- Operational metrics show how much work the data stewardship program is performing by counting things like how many data elements have been defined and owned, how many stewards have been trained, how often the data stewardship metadata is being used, and other measures of the program's efforts.

Rating your data stewardship maturity

- The maturity of your data stewardship effort measured in levels across a set of dimensions – is an important way to rate how the program is progressing and becoming more robust. The first step is to establish the levels and dimensions, most likely based on some maturity model available in the literature, then adjust them for your organization.
- The next step is to rate the current maturity levels and determine what the target levels should be. This exercise enables you to identify and prioritize areas of improvement so that you can get more out of the data stewardship program. Finally, you need to revisit the maturity on a periodic basis to determine the current maturity level and see if progress is being made.

Organization structure

- An organizational structure is crucial to successful data governance and data stewardship. This organization often has three levels, with the executive steering committee at the top, then the data governors in the middle, and the data stewards at the base.
- The organization is supported by various functions in IT, including analysts, developers, and technical data stewards. The organization is also supported and run by the DGPO, which handles the running of data governance board meetings and data stewardship council meetings.
- The DGPO also takes care of logistics, such as instituting a web portal, ensuring that all work is properly documented, making progress reports to executive and the data governance sponsor, and managing tools such as the issue log.

Summary

- Data stewardship is the operational part of data governance, and without data stewardship the best that can be hoped for in data governance is theory and perhaps some policies.
- Data stewardship and the data stewards are what make data governance a reality, with processes, procedures and data being understood, governed and improved in quality, and in general being treated like the enterprise asset is should be.
- There are many different kinds data stewards, such as business data stewards, technical data stewards, project data stewards, domain data stewards and operational data stewards. Each has a role to play, and data stewards need to work together to achieve the desired results.
- An organization with a robust and mature data stewardship effort can make decisions and answer questions about data efficiently and accurately, as well as being much more successful in improving data quality, executing on master data management, managing metadata, having more successful projects with more predictable timelines, and having the business take responsibility and accountability for the data they own.
- With successful data stewardship, championship of proper data management will emerge across the enterprise, from frontline data input specialists to data analysts to developers, spreading into management and to the executive levels, as company management see how accountable and responsible data management leads to better business decisions, better customer satisfaction, and better business opportunities.