2
The IBM Data Governance Unified Process

The benefits of a commitment to a comprehensive enterprise Data Governance initiative are many and varied, and so are the challenges to achieving strong Data Governance.

Many enterprises have requested a process manual that lays out the steps to implement a Data Governance program. Obviously, every enterprise will implement Data Governance differently, mainly due to differing business objectives. Some enterprises might focus on data quality, others on customer-centricity, and still others on ensuring the privacy of sensitive customer data. Some organizations will embrace a formal Data Governance program, while others will want to implement something that is more lightweight and tactical.

Regardless of these details, every organization should perform certain steps to govern its data. The IBM Data Governance Unified Process shown in Figure 2.1 maps out these 14 major steps (ten required steps and four optional tracks), along with the associated IBM software tools and best practices to support an effective Data Governance program.
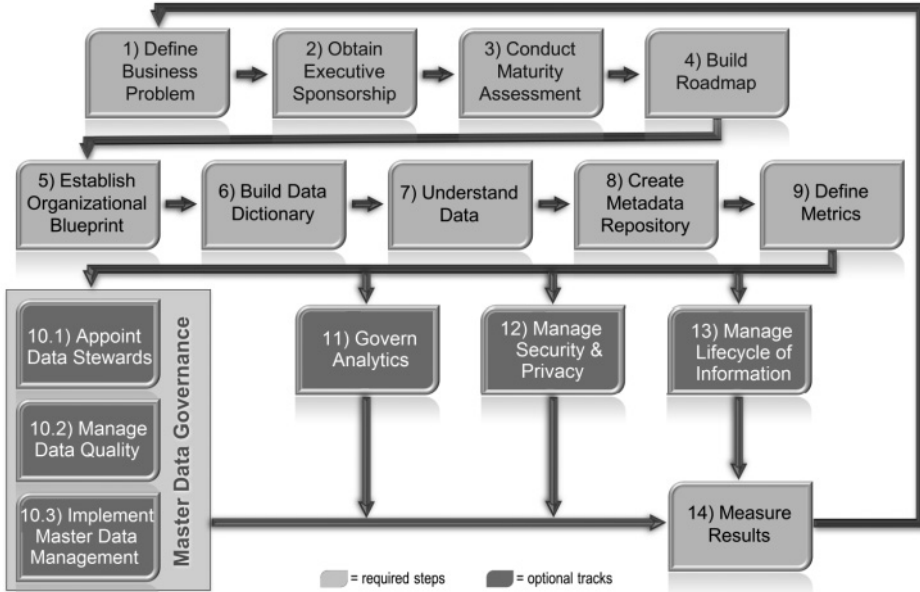
Figure 2.1: An overview of the IBM Data Governance Unified Process.

The ten required steps are necessary to lay the foundations for an effective Data Governance program. An enterprise will then select one or more of the four optional tracks, namely Master Data Governance, Analytics Governance, Security and Privacy, and Information Lifecycle Governance. Finally, the Data Governance Unified Process needs to be measured, and the results conveyed to executive sponsors, on a regular basis.

Let's walk through the steps in the figure in further detail:

1. *Define the business problem.*

   The main reason that Data Governance programs fail is that they do not identify a tangible business problem. It is imperative that the organization defines the initial scope of the Data Governance program around a specific business problem, such as a failed audit, a data breach, or the need for improved data quality for risk-management purposes. Once the Data Governance program begins to tackle the identified business problems, it will receive support from the business functions to extend its scope to additional areas.

2. *Obtain executive sponsorship.*

   It is important to establish sponsorship from key IT and business executives for the Data Governance program. The best way to obtain this sponsorship is to establish value in terms of a business case and "quick hits." For example, the business case might be focused on householding and name-matching, to improve the quality of data to support a customer-centricity program.

   As with any important program, the organization needs to appoint an overall owner of Data Governance. Organizations have historically identified the chief information security officer as the owner of Data Governance. Today, however, the ownership of Data Governance tends to reside within the CIO's office, in either the business intelligence or data architecture area. Data Governance leadership might also reside with the chief risk officer, especially in banks. A growing number of enterprises are staffing Data Governance roles on a full-time basis, with titles such as "data steward" indicating the importance of treating data as an enterprise asset. Regardless of title, the responsibility assigned to this role must be high enough in the executive ranks to ensure that the Data Governance program drives meaningful change.

3. *Conduct a maturity assessment.*

   Every organization needs to conduct an assessment of its Data Governance maturity, preferably on an annual basis. The IBM Data Governance Council has developed a maturity model based on 11 categories (discussed in Chapter 5), such as "Data Risk Management and Compliance," " Value Creation," and "Stewardship." The Data Governance organization needs to assess the organization's current level of maturity (current state) and the desired future level of maturity (future state), which is typically 12 to 18 months out. This duration must be long enough to produce results, yet short enough to ensure continued buy-in from key stakeholders.

4. *Build a roadmap.*

   The Data Governance organization needs to develop a roadmap to bridge the gap between the current state and the desired future state for the 11 categories of Data Governance maturity. For example, the Data Governance organization might review the maturity gap for Stewardship and determine that the enterprise needs to appoint data stewards to focus

on targeted subject areas such as customer, vendor, and product. The Data Governance program also needs to include "quick hits"—areas where the initiative can drive near-term business value.

5. *Establish an organizational blueprint.*

   The Data Governance organization needs to build a charter to govern its operations, and to ensure that it has enough authority to act as a tiebreaker in critical situations. Data Governance organizations operate best in a three-tier format. The top tier is the Data Governance council, which consists of the key functional and business leaders who rely on data as an enterprise asset. The middle tier is the Data Governance working group, which consists of middle managers who meet more frequently. The final tier consists of the data stewardship community, which is responsible for the quality of the data on a day-to-day basis.

6. *Build the data dictionary.*

   Effective management of business terms can help ensure that the same descriptive language applies throughout the organization. A data dictionary or business glossary is a repository with definitions of key terms. It is used to gain consistency and agreement between the technical and business sides of an organization. For example, what is the definition of a "customer"? Is a customer someone who has made a purchase, or someone who is considering a purchase? Is a former employee still categorized as an "employee"? Are the terms "partner" and "reseller" synonymous? These questions can be answered by building a common data dictionary. Once implemented, the data dictionary can span the organization to ensure that business terms are tied via metadata to technical terms, and that the organization has a single, common understanding.

7. *Understand the data.*

   Someone once said, "You cannot govern what you do not first understand." Few applications stand alone today. Rather, they are made up of systems, and "systems of systems," with applications and databases strewn all over the enterprise, yet integrated, or at least interrelated. The relational database model actually makes matters worse by fragmenting business entities for storage. But how is everything related? The Data Governance team needs to discover the critical data relationships across the enterprise.

Data discovery may include simple and hard-to-find relationships, as well as the locations of sensitive data within the enterprise's IT systems.

8. *Create a metadata repository.*

Metadata is data about data. It is information regarding the characteristics of any data artifact, such as its technical name, business name, location, perceived importance, and relationships to other data artifacts in the enterprise. The Data Governance program will generate a lot of business metadata from the data dictionary and a lot of technical metadata during the discovery phase. This metadata needs to be stored in a repository so that it can be shared and leveraged across multiple projects.

9. *Define metrics.*

Data Governance needs to have robust metrics to measure and track progress. The Data Governance team must recognize that when you measure something, performance improves. As a result, the Data Governance team must pick a few Key Performance Indicators (KPIs) to measure the ongoing performance of the program. For example, a bank will want to assess the overall credit exposure by industry. In that case, the Data Governance program might select the percentage of null Standard Industry Classification (SIC) codes as a KPI, to track the quality of risk management information.

These are the first nine required steps. The final required step is discussed later in this chapter. The enterprise also needs to select at least one of the four optional Data Governance tracks (Master Data Governance, Analytics Governance, Security and Privacy, and Information Lifecycle Governance).

Let's select the Master Data Governance optional track and walk through the application of its required sub-steps. The organization will need to ensure that the business problem (such as customer-centricity) is clearly articulated, and that executive sponsors are identified in the business and in IT. The organization will conduct a short Data Governance maturity assessment and define a roadmap. There needs to be some level of Data Governance organization to align the business and IT, to ensure near-term benefits. Business terms such as "customer" need to be clearly defined, especially if "customer" is one of the master data domains. The Data Governance organization needs to understand existing data sources and critical data elements. The business definitions, and the technical metadata from the discovery process, need to be captured within a metadata repository. Finally, the Data Governance organization needs to establish KPIs,

such as a reduction in customer duplicates, to measure the ongoing performance of the Master Data Governance program.

The level of emphasis on the required steps will vary based on the optional tracks that have been selected for Data Governance. As an example, let's review how step 7 ("Understand the Data") might be applied differently, based on the optional track or tracks selected. The Master Data Governance track will involve understanding the critical data elements to facilitate the mapping of sources to targets. The Analytics Governance track will involve understanding the relationship between key reports and critical data elements. The Security and Privacy track will involve understanding the location of sensitive data. Finally, the Information Lifecycle Governance track will enable the enterprise to understand the location of business objects, such as customer, as a precursor to an archiving project.

We will discuss these topics in greater detail in subsequent chapters, so we will just cover a few sample questions and potential focus areas for the remainder of this chapter. Here is a short description of the optional tracks within the IBM Data Governance Unified Process:

10. *Govern master data.*

    The most valuable information within an enterprise—the business-critical data about customers, products, materials, vendors, and accounts—is commonly known as *master data*. Despite its importance, master data is often replicated and scattered across business processes, systems, and applications throughout the enterprise. Governing master data is an ongoing practice, whereby business leaders define the principles, policies, processes, business rules, and metrics for achieving business objectives, by managing the quality of their master data.

    Challenges regarding master data tend to bedevil most organizations, but it is not always easy to get the right level of business sponsorship to fix the root cause of the issues. As a result, it is important to justify an investment in a master data initiative. For example, consider an organization such as a bank, which is sending multiple pieces of mail to the same household. This bank can establish a quick return on investment by cleansing its customer data to create a single view of "household." The bottom line is that the vast majority of Data Governance programs deal with issues around data stewardship, data quality, master data, and compliance.

11. *Govern analytics.*

    Enterprises have invested huge sums of money to build data warehouses to gain competitive insight. However, these investments have not always yielded results; as a consequence, businesses are increasingly scrutinizing their investments in analytics. We define the "Analytics Governance" track as the setting of policies and procedures to better align business users with the investments in analytic infrastructure. Data Governance organizations need to ask the following questions:

    ❑ How many users do we have for our data, by business area?

    ❑ How many reports do we create, by business area?

    ❑ Do the users derive value from these reports?

    ❑ How many report executions do we have per month?

    ❑ How long does it take to produce a new report?

    ❑ What is the cost of producing a new report?

    ❑ Can we train the users to produce their own reports?

    Many organizations will want to set up a Business Intelligence Competency Center (BICC) to educate users, evangelize business intelligence, and develop reports.

12. *Manage security and privacy.*

    Data Governance leaders, especially those who report in to the chief information security officer, often have to deal with issues around data security and privacy. Some of the common data security and privacy challenges include the following:

    ❑ Where is our sensitive data?

    ❑ Has the organization masked its sensitive data in non-production environments (development, testing, and training) to comply with privacy regulations?

    ❑ Are database audit controls in place to prevent privileged users, such as DBAs, from accessing private data, such as employee salaries and customer lists?

13. *Govern the information lifecycle.*

    Unstructured content makes up more than 80 percent of the data within the typical enterprise. As organizations move from Data Governance to Information Governance, they start to consider the governance of this unstructured content.

    The lifecycle of information starts with data creation and ends with its removal from production, and shredding from existence. Data Governance organizations have to deal with the following issues regarding the lifecycle of information:

    ❑ What is our policy regarding digitizing paper documents?

    ❑ What is our records management policy for paper documents, electronic documents, and email? (In other words, which documents do we maintain as records? For how long?)

    ❑ How do we archive structured data to reduce storage costs and improve performance?

    ❑ How do we bring structured and unstructured data together under a common framework of policies and management?

After these optional tracks, there is one more required step at the end of the Data Governance Unified Process:

14. *Measure the results.*

    Data Governance organizations must ensure continuous improvement by constantly monitoring metrics. In step 9, the Data Governance team sets up the metrics. In this step, the Data Governance team reports on the progress against those metrics to senior stakeholders from IT and the business.

The entire Data Governance Unified Process needs to operate as a continuous loop. The process needs to measure results and loop back to the executive sponsors for the continued endorsement of the Data Governance program.