TDWI Data Governance Principles and Practices

Managing Data as an Asset
Previews of TDWI course books offer an opportunity to see the quality of our material and help you to select the courses that best fit your needs. The previews cannot be printed.

TDWI strives to provide course books that are content-rich and that serve as useful reference documents after a class has ended.

This preview shows selected pages that are representative of the entire course book; pages are not consecutive. The page numbers shown at the bottom of each page indicate their actual position in the course book. All table-of-contents pages are included to illustrate all of the topics covered by the course.
<table>
<thead>
<tr>
<th>Module</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 1</td>
<td>Data Governance Concepts</td>
<td>1-1</td>
</tr>
<tr>
<td>Module 2</td>
<td>Data Governance Programs</td>
<td>2-1</td>
</tr>
<tr>
<td>Module 3</td>
<td>Applied Data Governance</td>
<td>3-1</td>
</tr>
<tr>
<td>Module 4</td>
<td>People and Data Governance</td>
<td>4-1</td>
</tr>
<tr>
<td>Module 5</td>
<td>Building and Evolving a Data Governance</td>
<td>5-1</td>
</tr>
<tr>
<td></td>
<td>Program</td>
<td></td>
</tr>
<tr>
<td>Module 6</td>
<td>Summary and Conclusion</td>
<td>6-1</td>
</tr>
<tr>
<td>Appendix A</td>
<td>Bibliography and References</td>
<td>A-1</td>
</tr>
<tr>
<td>Appendix B</td>
<td>Example Tools and Templates</td>
<td>B-1</td>
</tr>
</tbody>
</table>
You will learn:

✓ Definitions and dimensions of data governance
✓ Practices and techniques for building and operating a data governance program
✓ The roles, skills, and disciplines essential to data governance
✓ The importance of data stewardship to data governance success
✓ Activities, issues, and options when building a data governance program
✓ How to create the right balance of authoritarian and collaborative governance for the needs of self-service, agile development, big data, cloud services, and other modern practices
Module 1
Data Governance Concepts

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defining Data Governance</td>
<td>1-2</td>
</tr>
<tr>
<td>Dimensions of Data Governance</td>
<td>1-8</td>
</tr>
<tr>
<td>Data Governance Challenges</td>
<td>1-20</td>
</tr>
<tr>
<td>Multiple Approaches</td>
<td>1-34</td>
</tr>
</tbody>
</table>
Defining Data Governance

Governance Defined

Data governance is an emerging, cross-functional management program that treats data as an enterprise asset; a collection of corporate policies, standards, processes, people and technology essential to managing critical data to a set of goals.  

Maria Villar & Theresa Kushner

Data governance is the organization and implementation of policies, procedures, structure, roles, and responsibilities which outline and enforce rules of engagement, decision rights, and accountabilities for the effective management of information assets.  

John Ladley & Danette McGilvray


Today’s understanding of data governance should include:

- People who affect and are affected by data governance
- Data governance culture
- Collaboration as well as control
Defining Data Governance

Governance Defined

**MANAGING A VALUABLE ASSET**

Maria Villar and Theresa Kushner defined data governance as a “management program that treats data as an enterprise asset: a collection of corporate policies, standards, processes, people, and technology …”¹

John Ladley and Danette McGilvray defined data governance as “policies, procedures, structure, roles and responsibilities which outline and enforce rules of engagement, decision rights, and accountabilities for the effective management of information assets.”²

For years, data governance has been defined as an asset management practice with attention to data-related policies.

**EVOLUTION OF DATA GOVERNANCE**

With the rise in new data management technologies, innovative development methods, and increases in data volume, modern enterprises need to minimize bureaucracy without compromising effectiveness.

When building on previous definitions, two important areas must be adjusted:

- Add the human element. People govern, and they govern the behaviors of other people, not of the data itself. Human factors are important to achieve buy-in, participation, engagement, and adoption. Data governance culture is perhaps more important than data governance processes.

- Governance is more than control. Capabilities and services with a less enforcement-oriented approach enable people and projects to more easily apply good data management practices.

---

¹ Source: *Data Governance Fundamentals*, www.elearningcurve.com. Villar & Kushner are authors of *Managing Your Business Data*.

² Source: *Executing Data Quality Projects* by McGilvray. Ladley is a well-known EIM consultant.
Data Governance Challenges

What Data to Govern?
Data Governance Challenges

What Data to Govern?

**SCOPE OF DATA**

One of the big challenges when starting or improving a data governance program is to determine the scope of data to be governed. Every organization has lots of data. It is present in enterprise systems and databases, data warehouses, decision support databases, departmental systems, shadow systems, end-user databases, spreadsheets, and more. Data comes from financial transactions, online interactions, sensors, customer service reps, vendors and partners, etc. Furthermore, the data encompasses many subjects—customers, products, orders, accounts, employees, etc.—and each subject has different needs and considerations for quality, security, and compliance.

The combination of abundant and often redundant data with many data subjects makes scope of data a complex set of questions. Looking at a single subject—customer, for example—the questions include:

- Do you need to govern customer data?
- What are the motivations for governing customer data—quality, security, or compliance?
- Where are all of the places that customer data exists, including end-user databases and spreadsheets?
- For each location where customer data exists, is data governance necessary? Is it practical?

These are not easy questions to answer, and the answers will vary from one business to another. Consider, for example, the implications of customer data in a shared spreadsheet in a private cloud. If the business is healthcare and the customer data is really patient data, the security and compliance issues are significant. If the business is media services, this may be a lower risk scenario. It may be necessary to govern spreadsheets in one instance and impractical to do so in another.

**SOME GUIDELINES**

If you’re just getting started with governance it is wise to start with a limited scope—one or a few subjects with a high degree of cross-functional business activity. The Data Governance Institute advises to govern “as little as will help you meet your goals.” Also consider for each subject and for each kind of database the level of business interest and participation that you can expect. What level of support and sponsorship is realistic? What level of resistance is likely?
Module 2
Data Governance Programs

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Governance and Management</td>
<td>2-2</td>
</tr>
<tr>
<td>Data Management Processes</td>
<td>2-4</td>
</tr>
<tr>
<td>Program Development Processes</td>
<td>2-6</td>
</tr>
<tr>
<td>Program Operation Processes</td>
<td>2-14</td>
</tr>
<tr>
<td>Program Sustaining Processes</td>
<td>2-22</td>
</tr>
<tr>
<td>Program Growth Processes</td>
<td>2-26</td>
</tr>
</tbody>
</table>
Governance and Management

Managing a Governance Program
Governance and Management
Managing a Governance Program

LEVELS OF GOVERNANCE
This module will address all the pieces needed for an exhaustive data governance program. A more hierarchical data governance program may carefully document how each of these capabilities will be fulfilled before getting started; a more collaborative governance program may focus on data management and develop the other capabilities gradually.

However they are executed, the capabilities, goals, and responsibilities that follow are important for the smooth function and effective operation of a program of any size or structure.

MANAGING A GOVERNANCE PROGRAM
A complete data governance program requires program management across four broad responsibilities:

• **Develop** and establish a new governance program at its inception
• **Operate** data governance on a day-to-day basis
• **Sustain** the program as issues arise and scope evolves
• **Grow** governance capabilities and data management maturity
Program Growth Processes

Change Management
Program Growth Processes

Change Management

GROWTH BY ACTIVE CHANGE MANAGEMENT

Some changes are predictable and others are not. Some are within our control and others are not. The scope and priorities management previously described are necessary to react to unanticipated change. However, not all change management is reactive. We can often anticipate change, especially changing business needs, stakeholders, data, systems, and technology.

To actively manage change, keep these principles in mind:

- Planned change depends on defined strategy and architecture.
- When change is anticipated, it can be managed proactively.
- When externally driven change occurs unexpectedly, it is often possible to anticipate and proactively manage internal changes.
- Change never happens without issues arising. Change management needs a good issue resolution process.
- Organizational change is highly dependent on good communication. Change management needs a good communications plan.
- Desired change (improved data management practices, improved data quality, information management maturity, etc.) doesn’t happen by accident. It only occurs when change is managed.
Module 3
Applied Data Governance

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Governance Through the Information Lifecycle</td>
<td>3-2</td>
</tr>
<tr>
<td>Agile Data Governance</td>
<td>3-4</td>
</tr>
<tr>
<td>Big Data Governance</td>
<td>3-22</td>
</tr>
<tr>
<td>Cloud Data Governance</td>
<td>3-32</td>
</tr>
<tr>
<td>Self-Service Data Governance</td>
<td>3-38</td>
</tr>
</tbody>
</table>
Governance Through the Information Lifecycle
Positioning Processes

What new factors are reshaping the governance lifecycle?

- Capture
- Analyze
- Maintain
- Use
- Distribute
- Preserve/Dispose
Governance Through the Information Lifecycle
Positioning Processes

END-TO-END GOVERNANCE

Data has a finite lifecycle that begins with capture and ends with disposal. Data governance must attend to all stages of the lifecycle with policies, processes, responsibilities, and accountabilities related to each stage.

*Capture* includes the acquisition, creation, and receipt of data, e.g., recording new entities such as customers and products, recording business transactions, and receipt of data from external sources.

*Distribution* of data internally in preparation for use. Privacy and security policies are paramount when managing data distribution. This phase includes many standard data management practices.

During *analysis*, data passes through analytics processes to create new values, potentially with more or different value to the enterprise. Not all data goes through an analysis phase, but it is becoming more common.

*Use* logically follows distribution and analysis, and at this point, data is applied to business processes, activities, and decisions. Appropriate use—the right data for the right purpose—is an important element of data governance and data value management.

*Maintenance* includes filing, updating, indexing, and retrieval of data and information. The data might be published externally as well.

*Archiving* data handles data that is infrequently accessed, that has declined in value, or that has met its designated retention period. *Disposal* of data may be necessary to meet regulatory requirements of privacy and confidentiality. Conversely, regulatory requirements sometimes mandate that data be preserved.

NEW PROCESSES

Changes in both the technical management of data and the human use of data are creating the need for new kinds of data governance. Agile methods are increasing the desire for flexible, fast processes. The adoption of big data means that more data, and more kinds of data, must be governed faster. Cloud platforms create a new venue for governance, and self-service practices increase governance challenges by increasing the need for access.
Agile Data Governance

Agile Process and Values

Individuals and interactions over processes and tools
Working software over comprehensive documentation
Customer collaboration over contract negotiation
Responding to change over following a plan
Agile Data Governance

Agile Processes and Values

AGILE SYSTEMS DEVELOPMENT

Agile development is based on the principle of discovering requirements through active collaboration with business subject experts. Agile projects are short in duration and fast moving, with each project delivering a small set of functionalities and business capabilities. Several projects are typically required to build a robust system. Projects are executed through multiple iterations until a release-ready product is built. Iterations include daily planning and direction for a continuous build process. When a release-ready product is completed, the release activity begins, including planning, documentation and formalization, production implementation, etc. Releases and iterations achieve cohesion and continuity by working within a strategic structure of shared vision and goals.

THE AGILE MANIFESTO

Several of the pioneers and leading practitioners of agile development (too many to list here) collectively authored the “Manifesto for Agile Software Development.” Following these principles will encourage quick and nimble systems development whether it’s called “agile” or not.

“We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:

• Individuals and interactions over processes and tools
• Working software over comprehensive documentation
• Customer collaboration over contract negotiation
• Responding to change over following a plan

That is, while there is value in the items on the right, we value the items on the left more.”

GOVERNANCE: FRIEND OR FOE?

It is a common belief that agile and governance are in conflict—that the processes of governance are too burdensome and become barriers to agile development. In reality, agile and governance can work together quite effectively, but it requires a change in mindset. We need to think differently about agile teams and about how we govern—making the work of governance part of the agile process and focusing governance on value produced instead of processes followed.

1 http://agilemanifesto.org
# Module 4

**People and Data Governance**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positioning Data Governance in the Enterprise</td>
<td>4-2</td>
</tr>
<tr>
<td>Governance and Management Practices</td>
<td>4-4</td>
</tr>
<tr>
<td>Data Governance Roles</td>
<td>4-10</td>
</tr>
<tr>
<td>Stewardship Concepts</td>
<td>4-22</td>
</tr>
<tr>
<td>Stewardship Organizations</td>
<td>4-26</td>
</tr>
<tr>
<td>Stewardship Skills and Knowledge</td>
<td>4-32</td>
</tr>
<tr>
<td>Data Curation</td>
<td>4-40</td>
</tr>
</tbody>
</table>
Governance and Management Practices
Responsibility, Authority, and Accountability
# Governance and Management Practices

## Responsibility, Authority, and Accountability

**PEOPLE GOVERN**

Data governance is something that is done by people. Many people in various roles interact to govern data. Every role has designated responsibility, authority, and accountability. Even in a collaborative governance program that does not embrace formal roles, the tasks that would fall to traditional roles need to be accomplished. Responsibility, authority, and accountability are the cornerstones of governance.

**RESPONSIBILITY**

*Responsibility* is the obligation incurred by an individual in a specific role to perform the duties of that role. The individual is obligated to take actions and produce results that affect the organization’s assets.

**AUTHORITY**

*Authority* is the power granted to an individual in a specific role to make decisions and direct others to follow those decisions.

**ACCOUNTABILITY**

*Accountability* is the individual liability created by use of authority. It is a condition of being fully answerable for results and achievement of goals.

**ALIGNMENT**

A governance organization (or any organization) works well when responsibility, authority, and accountability are well *aligned*. The authority granted to a role must be aligned with the responsibilities of that role. The accountability of the role must correspond with the scope and level of authority.

Accountability is the logical consequence of responsibility and authority. Gaps and inconsistencies between the three elements are certain to cause dysfunction in the organization.

**DELEGATION**

*Delegation* is an additional consideration to keep in mind when defining governance roles. A particular role may be assigned a particular responsibility with or without the right to delegate that responsibility. When responsibility is delegated, the corresponding authority must also be delegated.

Responsibility and authority may be delegated. Accountability cannot be delegated—it remains with the delegator.
Data Governance Roles

Governance as Teamwork
Data Governance Roles

Governance as Teamwork

THE GOVERNANCE TEAM

A formal data governance team includes four roles as illustrated on the facing page:

- A data executive designated as the person who provides overall leadership of a data governance program
- Data owners responsible for access, distribution, retention, etc.
- Data stewards who facilitate consensus data definitions and foster sound data quality, data usage, and data security practices
- Data specialists such as data architects, data modelers, database developers, and database administrators who have custodial responsibility for data

THE DATA EXECUTIVE

The data executive role is handled differently in various organizations and data governance programs. In some instances, the data executive is a C-level position. Governance responsibilities might be covered by a chief data officer (CDO) role. The CIO or the CFO may be designated as having executive responsibility for data. Security- and compliance-driven programs may designate a senior security or compliance officer.

In other instances, a data governance program manager is created as a senior management position that provides a bridge for engagement between director and CxO positions.

No matter what makes the most sense for your organization’s size, needs, governance goals, and culture, it is important to include executive perspective in the data governance program.
# Module 5

Building and Evolving a Data Governance Program

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balanced Data Governance</td>
<td>5-2</td>
</tr>
<tr>
<td>Growing and Evolving Governance</td>
<td>5-14</td>
</tr>
<tr>
<td>Technology for Governance</td>
<td>5-18</td>
</tr>
</tbody>
</table>
Balanced Data Governance
Continuum

Control | Collaboration

Decision Rights and Decision Models

Enforcement and Prevention

Hierarchies and Networks

Rigor and Agility
Balanced Data Governance

Continuum

**GOVERNANCE FIT FOR PURPOSE**

We’ve discussed how data governance can seem pulled in two directions: on one extreme, a highly centralized governance group handing down strict policies that are slow to change and keeping rigid control of data. On the other, a free-for-all resulting in privacy breaches and inconsistent, untrustworthy data analysis. Neither of these visions are good for a modern enterprise, but they are both caricatures.

Depending on internal culture and the needs of a specific department or project, data governance needs vary. For example, management of loan approvals is a high-risk, mission-critical activity that deals with sensitive information. It must be subject to carefully managed decision rights, and will be the focus of centralized stewardship and governance. On the other hand, departmental analysts exploring market trends by combining historical transactional data with external market data requires more flexibility and deserves more autonomy.
Technology for Governance

Applications

- Visualization
- Reporting
- Dashboards
- Scorecards
- OLAP
- Analytics
- Machine Learning

- Pipelines
- Workflows
- Lineage
- Automation
- Auditability

- Cleansing
- Selection/Sampling
- Transformation
- Aggregation
- Integration

- Metadata Management
  - Data Catalog
  - Data Storefront
  - Data Models

- Acquiring
- Documenting
- Processing
- Delivering
- Using

Access
- Interfaces
- Extraction
- Profiling
Technology for Governance

Applications

TOOLS FOR GOVERNANCE

Data governance, just like any other important business function, needs technology to be successful. The market offers a handful of software tool suites for data governance. However, a governance tool suite isn’t always necessary; standalone tools may be selected for areas with the greatest need or based on a “best of breed” approach. Governance tools are also included in products that support data modeling, integration, cleansing, preparation, or cataloging. Alternatively, many organizations choose to develop their own solutions internally.

Data governance tools should enable functions across the following areas:

- **Acquiring data**: governing access, profiling, extraction
- **Documenting data**: metadata management, data cataloging, data storefronts, data models
- **Processing data**: governing cleansing, selection and sampling, transformation, integration and blending
- **Delivering data**: governing data pipelines and workflows, tracking lineage, automation, providing an audit trail
- **Using data**: governing visualizations, reports, dashboards, scorecards, OLAP, analytics