

Enterprise Content Management and Data Governance Policies and Procedures Manual

Shamnad Mohamed Shaffi

Colorado Technical University

Abstract: Enterprise Content Management (ECM) and Data Governance are essential for organizations to manage, secure, and optimize data assets efficiently. This paper presents a structured framework for implementing ECM and data governance policies using an imaginary company, Teleware, as a case study. The study explores common challenges faced by enterprises, including data integration, quality, accessibility, and regulatory compliance. The proposed framework includes an assessment of the existing information infrastructure, identification of regulatory requirements, and enhancement of content management processes. A phased implementation approach is introduced, incorporating best practices in metadata management, data security, and risk mitigation. The adoption of a next - generation data governance platform ensures improved data quality, workflow efficiency, and enterprise - wide accessibility. Agile methodologies are leveraged to streamline policy execution, ensuring adaptability to evolving business needs. This framework serves as a strategic guide for organizations aiming to establish a data - driven culture, enhance compliance, and optimize decision - making through effective content and data governance

Keywords: Enterprise Content Management, Data Governance, Data Security, Metadata Management, Risk Mitigation, Compliance, Workflow Optimization, Agile Methodologies, Data - Driven Decision – Making

1. Project Outline and Requirements

1.1 Content Requirements

Teleware company has recorded tremendous growth in the last several quarters, and the product and Technology team has implemented many modernization projects and critical capabilities to drive the business. As the Company operates in many business lines like prepaid, postpaid, and wholesale, there are different billing systems involved and many different platforms for the supply chain management. The data analytics platform that drives the strategic reporting for finance, supply chain, and marketing receives a variety of data from many completely different sources in structured, semi - structured, and unstructured formats. The data from these different platforms are integrated into a centralized data warehouse and Datamart's are build based on different business requirements.

A survey was recently conducted across the Enterprise, and as part of that initiative, 50 stakeholders from Marketing, Supply Chain Management, Finance, and Sales were interviewed. The management team collected the information about the business users' pain points, workflows, how they interact with the different groups, and their vision of an ideal data platform and working relationship with the technology team. The top 5 pain points across all users related to data integration, the lack of a data dictionary, the need for more significant subject matter expertise, the need for higher data quality standards, and simplified access to enterprise data.

The Company currently lacks a centralized repository and process to manage their data assets. One critical vision of the technology team is to build a content management framework to bring together the people, processes, and technologies needed to create a data - driven culture. The following are the key objectives of this project –

- Conduct an initial assessment and evaluate the existing enterprise information infrastructure resources and content management process.
- Deploy the Next Generation Data Governance Platform comprised of three products that will help us discover, understand, manage, govern, share, secure, and use enterprise data

1.2 Content Categories

According to the initial assessment, there are more than 20 source applications ingesting data into data analytics platforms in different data formats, mainly structured and unstructured.

1) Unstructured Content

Unstructured Data is Data with no predefined structure and can be stored as different files, including image files, PDFs, and text documents (Word, .txt, PPT, videos, etc.). Also, natural language presented in structured formats is considered unstructured data (includes chat, messages, blogs, etc. stored in databases/tables). Unstructured Data has no predefined table or data structure and usually has little long - term business value. "point in time" file or document. The following are the key observations from the initial analysis -

- a) Unstructured Data usually has little long - term business value.
- b) An individual PPT, Word doc, image, etc. usually have a limited purpose and often does not provide enterprise value.
- c) There are far more unstructured data owners than structured data owners.
 - Unstructured data owners can be any business user that creates PPTs, Word documents, etc.
 - Unstructured data owners are usually less technical than their structured data counterparts.
- d) Unstructured data can be found in structured data environments.

- For example, a database may have a field that contains chat conversations with support agents. The chat data is unstructured but may reside in a structured environment and format.

2) Structured Content

Structured data refers to fixed field data, typically represented by schema, and can be represented by relational database rows and columns or spreadsheets. Structured data can easily be entered, captured, and maintained in a relational RDBMS database application like Teradata, Oracle. The data can be easily queried using Structured Query Language (SQL) and analyzed by users. As per the analysis conducted, most source systems like Supply chain management, billing systems, etc. are capturing data in a structured format.

1.3 Associated Regulations

The Company also needs to ensure that the Enterprise meets the SOX compliance policies and other State's Privacy compliance requirements in the future - The right to access personal information, the right to stop the sale of personal information, and the right to request deletion, right to know about personal information in our care. Another key objective of the project is to establish a Data Governance process around new application onboarding enterprise - wide for privacy compliance.

1) Data Protection

The team should build a process framework for how to apply security measures to ensure that enterprise information is appropriately handled and protected. The information guidelines should provide overall guidance for the consistent application of security measures for managing data information. The following are the key assessments performed.

- Get an understanding of the encryption requirements and the process of encryption.
- Identify systems where encryption and decryption of data are needed.
- Review the Security Design Patterns and ensure tools to perform encryption and decryption are in place.
- Provide a Threat Model for the solution.

2) Implementation Life Cycle

A Software Development Lifecycle (SDLC) Framework is formally documented, approved, and maintained. The SDLC Framework includes planning activities, high - level system design (HLSD), detailed design, build, test, and deployment phases. The team has decided to adopt an Agile methodology that focuses on continuous delivery and change management for implementing the project. The following are the different stages of implementing the enterprise solution.

a) Requirements

Business stakeholders should formally define the business requirements, functional requirements, associated risks and controls, and roles and responsibilities. Any changes to the business or technical requirements are documented and approved by the appropriate business or IT stakeholders.

b) Planning

Once the requirements are baselined, the project team needs to conduct an initial requirements assessment, clarify if any open questions with business stakeholders and develop the project plan and timelines. Prioritizing the tasks, creating user stories, set goals and measures are defined during the planning.

c) Development

Once the user stories and priorities are defined, the stories will be taken by the development team based on their bandwidth for application development. The development team will be working closely with Architects, quality assessment team on creating test cases, and ensuring that the application developed is aligned with the architecture and delivers the most optimal code.

d) Review

The code developed will be tested in different testing and review phases, mainly unit testing to evaluate the functional testing, regression testing to ensure that the change is not breaking existing functionality, and performance testing to ensure that the application is scalable. Review processes are performed on each of these testing stages and the final approval request needs to be submitted before implementation which needs to be approved by the leadership team.

e) Implementation

The implementation plan includes identifying the best window to deploy the changes without or minimal impacting operations and business users. The plan also includes the backup plan to revert the changes in case of any failure during the implementation. The team also should send out an advance notification to inform all concerned stakeholders about the implementation.

f) Follow - up

Once the solution is implemented, it is vital to ensure that the project fixing any existing tech debt, if yes then please provide a brief description of tech debt and the proposed solution to fix it. Ensure end - user training and getting user feedback is considered in the solution, as needed.

2. Information Infrastructure Evaluation

The vision of Teleware Communications is to bring together the people, practices, and technologies needed to build a data - driven culture. Knowledge and information are among the most valuable assets any organization possesses (Stewart, 2012). The roadmap is to deploy an Enterprise Governance platform comprised of three products that will help us discover, understand, manage, govern, share, secure, and use enterprise data. The content has become so voluminous and diverse in its forms and how it comes into the Enterprise, that pretty much every organization experiences some level of content related dysfunction (Stewart, 2012). The evaluation approach varies by type - source, a system of record, analytical, middleware, API, etc. The assessment should include the following -

- Review the existing enterprise Contents, Policy, Standards, Procedures, Guidelines, Controls.

- Review current Data Governance Program, Standards, Stakeholders, Policy Concept, and Plan.
- Understand data governance requirements and expectations.
- Identify and select Technology, Business, and Compliance owner for each standard.
- Engage Key Stakeholders from across Enterprise to understand the current state.

The following are the components of evaluation of ECM system architecture (MUSE, n. d)

- Capture
- Management
- Storage
- Preservation
- Delivery

1) *Evaluation of Content Capture.*

The below diagram outlines the different source systems at Teleware communications Ltd, the source applications/platforms being used, the structure of data, and how its ingested into the target Datawarehouse. The finance datasets are being used for updating General Ledger and analytics on Account Receivable and Payables. Customer data and Customer Clickstream data is collected for marketing campaigns and customer acquisition. Order information is required to maintain the sale history, order lifecycle management, receipt generations, and care support.

2) *Evaluation of Content Management*

The Table 1 provides the summary of evaluation on the content management process performed at Teleware communications Ltd.

a) **Customer Data**

- Content: Customer MDM
- Comments: Customer data is not standardized across the enterprise, making it difficult to identify if the same customer interacts through multiple channels. Additionally, security controls to restrict sensitive and confidential information are lacking.

b) **Order Information**

- Content: SAP Supply Chain Management
- Comments: Order information from different sales channels (retail stores, websites) exists in different formats and is not consolidated into a common structure. Order lifecycle tracking is incomplete due to differences in channel processes.

c) **Finance Datasets**

- Content: Sales DB
- Comments: General Ledger (GL) information is captured on paper documents. Billing system data is not preserved, and no history is maintained in any data store. Versioning of data is incomplete, and security controls are inadequate for protecting sensitive financial data.

d) **HR and Time Keeping**

- Content: SharePoint
- Comments: Employee and timekeeping information are stored in SharePoint, but there is no process to accurately track employee schedules and working hours.

e) **Customer Clickstream Data**

- Content: Adobe Analytics

- Comments: Clickstream data is high in volume and unstructured. It is stored in a Hadoop data lake without a schema, making it difficult to extract value. Users are unfamiliar with data consumption from the data lake, and extracting the data is time - consuming.

3) *Evaluation of Content Storage*

At Teleware communication Ltd, the content from different source systems is collected and stored in different applications. The sales information from retail and non - retail channels is collected on the POS database, which is behind a firewall. The payments account receivable and payable information is stored in Amdocs payment gateway application and the order information is maintained at SAP supply chain application. Employee information, including the schedule and work hours are managed at SharePoint sites.

4) *Evaluation of Content Preservation*

The following are the summary of evaluation on the content preservation process performed at Teleware communications Ltd.

a) **Customer MDM**

- Storage Method: Data stored in the SQL database, with paper copies maintained as a backup in file cabinets.
- Comments: SQL database data is extracted into spreadsheets or Word documents for analysis. There is no standard template for report generation, leading to ad - hoc reporting.

b) **Order Information**

- Storage Method: Data stored in different databases; no backup or history is maintained.
- Comments: Order information is generated into Excel spreadsheets, PDFs, and Word documents. No standard template or automated end - to - end process is in place.

c) **Sales Database**

- Storage Method: Data stored in the SQL database, with paper copies maintained as a backup in file cabinets.
- Comments: Data is generated into stylesheets but lacks visualization tools like MicroStrategy, Tableau, etc.

d) **SharePoint Site**

- Storage Method: Data stored in SharePoint sites, accessible and editable by anyone with access.
- Comments: Employee information from SharePoint is printed as PDFs and handed over to employees and contractors for review.

e) **Adobe Analytics**

- Storage Method: Data stored in the Hadoop data lake as collected from different channels.
- Comments: Data is rarely used due to the lack of technology to extract different data formats. Structured customer interaction data is extracted into Excel, but unstructured data (e. g., clickstream) remains largely unused

5) *Evaluation of Content Delivery*

The finance and Marketing team manually retrieves the customer and sales dataset directly from the database application using different SQL query tools available in the market like Toad, SQL assistant, etc. The users combine the financial result sets pulled from different systems manually as required. For order information, currently, there isn't a cross - departmental platform to track the different

Lifecycles of orders from receipt creation until fulfillment. Similarly, the employee information recorded in the SharePoint site can be accessed by anyone having access to the Company's intranet. Customer interaction data stored in Hadoop is accessed through a tool called Ambari, but it cannot connect to SQL databases to join the different datasets.

6) Evaluation of ECM Relationships

A Customer relationship management (CRM) solution has been implemented at Teleware Communications Ltd that helps to increase the sales by identifying the specific customer requirements, trends and potential needs, improving customer services, reducing customer churn, refining the sales processes, manage business relationships, etc.

However, the lack of quality standards and practices to ensure the integrity of the Company's data assets is a challenge throughout the Enterprise to adhere to standards ensuring the integrity of information. The lack of a common understanding of data across Customer, Marketing, Sales, and Order management is a challenge to bring the value out of combined datasets to increase the revenue.

3. Information Infrastructure Improvements

As part of improving the enterprise information infrastructure at Teleware communications Ltd, the management has explored options to manage the procedures and policies for information asset management. The improved strategies will help to bring together the people, practices, and technologies that will help us discover, understand, manage, govern, share, secure, and use enterprise data. In any process - improvement initiative, there must be a primary focus. Without a primary focus, it is easy to lose sight of the initiative, and *scope creep* can happen. Not everything can be changed at once; therefore, a hierarchy and plan of action must be developed (MUSE).

1) Content Management Process Improvement

As per the assessment conducted at Teleware communications Ltd, currently, there is no tool used to collect data and ingest into the enterprise database. The front - end client consultants use paper forms to collect sales information, which is being sent to a data entry specialist who updates the enterprise database with the information captured through paper forms. The process is time - consuming and not efficient as it can cause manual errors while entering data, which could impact the decision - making process in the Company.

The recommended improved process will remove the role of data entry specialist from the data capture process and enables the client - consultant to enter the data using an electronic device like workstations or laptops connected to a secured network to update into an enterprise database. The new process would be efficient, secure, and reduces the time significantly on the data capture process. One key benefit of this improved process is the availability of data in near real - time, which would help management to view the sales happening now and make any immediate strategic decisions.

2) Content Management Tools Improvement

Currently, the legacy data flow from the age of data warehousing is focused on processing and storing data. Linear ETL is entirely process - oriented, with Data simply acting as the object of processing. The different stakeholders worked with limited and controlled data sources consisting primarily of structured enterprise operational data stored in relational databases

The below diagram outlines Legacy Data Flow for Information Delivery – focus on processing and storing data.



The improved information supply chain increases the focus on data. The purpose of processing is to move data through the stages of the supply chain – data ingestion, data cataloging, data preparation, and data analysis.

- a) Data Ingestion: Data from different sources are ingested in batch, micro - batch, stream processing and should have the ability to onboard new datasets, The process should be able to ingest all types of data from a variety of sources, should have full visibility into data mappings and data flows and proactively detect and prevent data pipeline defects.
- b) Data Catalog: Implement a solution to collect and maintain metadata in a secure repository. This enables us to connect people with data (find needed data, evaluate and understand dataset contents, acquire datasets). Metadata is the core of a data catalog that enables data searching, identify compliance - sensitive data, data lineage, and uses of data.
- c) Prepare: This facilitates combining data from multiple sources and making data ready for analysis and reporting. Data transformation is the essence of data preparation by changing data to meet specific needs and goals and getting the right data in the right forms and formats for analytics – standardize; conform; cleanse; quality assure, de - duplicate, plus derive, append, aggregate, join, sort, pivot, filter, mask, sample, construct, etc. are a few steps in the data preparing process.
- d) Analyze: Analyzing involves finding patterns, trends, and anomalies in the data. This helps in seeking business understanding and insight from the Data, support for descriptive, discovery, diagnostic, predictive, and prescriptive analytics.
- e) Act: The process facilitates fostering a culture of data analysis and federated analytics across the Enterprise (share the information); continued support for BI and decision support beyond insight.

3) Data Governance Evaluation

Data Governance is a system of decision rights and accountabilities for information - related processes, executed according to agreed - upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods (Thomas, n. d). Data governance is an evolving, cross - functional management program that treats Data, both internal and external, as enterprise assets. It defines the rules and practices to manage critical data to a set of goals that may include any or all of quality, security, privacy, compliance,

integration, and information utility. Data governance encompasses standards, processes, people, and technology. It fosters the adoption of data management rules and practices without unnecessarily compromising agility or constraining the scope of governed data. Currently, Teleware communications lack a formal data governance strategy. The processes and procedures are not well documented, and the leadership team lacks oversight on how to implement a successful data governance strategy. The below section outlines the data governance evaluation conducted at Teleware communications Ltd.

4) Data Quality

Data facilitates the ability of management to plan and meet strategic goals set for the organization (MUSE). Per ISO 9000, data quality can be defined as the degree to which a set of characteristics of data fulfills requirements. A data quality dimension is a term describing an aspect of data that can be measured or assessed against defined standards in order to determine the quality of data. The six core dimensions of data quality are:

- Completeness
- Uniqueness
- Timeliness
- Validity
- Accuracy
- Consistency

The following are the summary of assessment on the data quality process performed on the Teleware Communications Ltd.

Strengths:

- Have minimal data quality standard on ABC (Audit, Balancing & Control - a measure of data availability).
- Information Technology personnel are aware of data collected in different source systems.
- The company has data validation on front - end applications.

Weaknesses:

- All data is not mapped, access - controlled, and managed.
- Identify and describe any cautions and warnings to be communicated to data consumers.
- Lack of profile data to estimate accuracy and completeness.
- The current process is not capable of capturing and assessing the quality of available metadata.
- Minimal and least data quality monitoring and reporting.

5) Data Management

Teleware communications Ltd doesn't have a defined data management and governance strategy. The data definition of entities is not standardized across departments that establish a common understanding and technical and business definition for critical data. Data retention and the archiving process is not defined, and the audit balance control and data validation process are completely implemented.

Strengths:

- Minimum data management policies to meet current needs.

- Has a record of processing that maintains a record of all assets used for data processing.
- User - defined data for managing business rules and processes.

Weaknesses:

- All data is not mapped, access - controlled, and managed.
- Identifying any regulatory constraints for data retention, defining aging, archival, and purging criteria.
- No automated process to discover, identify, and tag data across systems, apply regulatory rules and contractual obligations, assess risk, and act.

6) Data Policies Management

Teleware communications Ltd has a dedicated team that governs the ownership of the Company's information and sets forth rules and requirements for the use of information and communication resources. The team is responsible for creating the procedures to addresses the entire development lifecycle for information security policies, including policy document creation, intake, review, approval, publication, communication, exceptions, updates, sunseting, and archival.

Strengths:

- Has a data governance team that manages and maintains the current data governance policies.
- The approval process to grant access to enterprise applications.
- Role - based permission for applications.

Weaknesses:

- Lack of data governance documentation status and completion.
- Challenges in renewing existing policies and retiring obsolete policies.
- Lack of continuous monitoring of sensitive data against GDPR, CCPA, subject access requests, right - to - be - forgotten, data sharing agreements, internal data management policies, and more.

7) Business Process Management

Business Process management allows an organization to automate routine or repetitive tasks and turn them into multi - step workflows. At Teleware communications Ltd, each business process starts with a trigger event (i. e., a form is submitted to the Shared Services Team to request support on a SharePoint site) kicking off to lead the processor towards their end goal. When you automate a business process, you identify the series of tasks that must be done to complete any item in a workflow. Before automation, all those tasks and notifications were generated manually. The management team in Teleware communications Ltd benefits from automating the below processes.

- Streamline communication
- Enforce accountability
- Minimize costs and manual errors
- Easier to manage.

Strengths:

- Has a data governance team that manages and maintains the current data governance policies.

- The approval process is in place to grant access to enterprise applications.
- Role - based permissions are implemented for applications.

Weaknesses:

- Lack of data governance documentation status and completion.
- Challenges in renewing existing policies and retiring obsolete policies.
- Lack of continuous monitoring of sensitive data against GDPR, CCPA, subject access requests, right - to - be - forgotten, data sharing agreements, internal data management policies, and more

8) Risk Management

Enterprise Risk Management team in Teleware communications Ltd is a service and governance function to all areas of the Company and is responsible for developing, deploying, documenting, and managing scalable and sustainable enterprise risk management programs. This includes the Company's overall risk management program that enables it to understand and manage risks that can have a material effect on the Company's financial condition and results of operations.

Strengths:

- The company has advisors to the business regarding insurance and risk mitigation.
- The Risk Management Team manages company risks through various risk transfer and mitigation efforts, including insurance policy procurement, claims management, and policy/procedure development.

Weaknesses:

- All data is not mapped, access - controlled, and managed.
- Lack of confidence and usability in enterprise and performance issues.
- Identifying risks of breach, deterioration, and on - demand disclosure through litigation.
- Identifying regulatory constraints and other inhibitors of data sharing

4. Data Governance Improvements

Large organizations must make many decisions about their IT systems and the data that flows through them. Many of these decisions are made by IT Governance groups focusing on IT Portfolio Management issues, such as deciding when it's time to put in a new application, with all its new pipes and pumps and storage tanks (Thomas, n. d)

1) Data Quality

Based on the research on the current enterprise data quality, the key focus should be to create a robust process and policies to ensure the state of completeness, validity, consistency, timeliness, and accuracy that makes data appropriate for a specific use. The new process ensures that the Data is complete, standards - based, consistent, accurate, and time - stamped. The policies should be revised at regular intervals to ensure that the enterprise standards for data quality are always met.

2) Data Management

Based on the research on the current enterprise data quality, the key focus is to Enhance and expand enterprise data governance and data management. Establish common understanding and technical definitions for critical data. The data management process should include defining a process for the following -

- Data Definition and Standardization,
- Data Retention, Purge and archival
- Audit Balance Control and Data Validation

3) Data Policies Management

The primary focus on the data policies management is to establish, communicate, and foster adoption of new standards and practices for enterprise content management. The new process would improve service level with goal setting, measurement, and monitoring. A robust change management and documentation process should be created as part of the policy management. These policies will help improve standardized governance, operational efficiency, and drive towards a holistic enterprise compliance program.

4) Business Process Management

The improved business process management at Teleware Communications Ltd will drive a common understanding of business data and give a unified view of enterprise performance and compliance. Maintain a record of all assets used for data processing as well as a registry of processing purposes that can be mapped to data terms and policies.

5) Risk Management

Enterprise Risk Management team in Teleware Communications Ltd should implement a service and governance function to all areas of the Company and is responsible for developing, deploying, documenting, and managing scalable and sustainable enterprise risk management programs. This includes the Company's overall risk management program that enables it to understand and manage risks that can have a material effect on the Company's financial condition and results of operations.

References

- [1] Building an effective data governance framework. (2013). Retrieved from Building an effective data governance framework: <https://searchdatamanagement.techtarget.com/essentialguide/Building-an-effective-data-governance-framework>
- [2] Evaluation of Data Governance. (n. d.). Retrieved from MUSE: https://class.ctuonline.edu/_layouts/MUSEViewer/MUSE.aspx?mid=17555284
- [3] Evaluation of the ECM System. (n. d.). Retrieved from MUSE: https://class.ctuonline.edu/_layouts/MUSEViewer/MUSE.aspx?mid=17555218
- [4] Javanainen, M. (2016). ECM: Continuous Improvement and the Theory of Constraints. Retrieved from ECM: Continuous Improvement and the Theory of Constraints: <https://www.m-files.com/blog/ecm-continuous-improvement-theory-constraints/>

- [5] MUSE. (n. d.). Implementing Data Governance. Retrieved from MUSE: https://class.ctuonline.edu/_layouts/MUSEViewer/Asset.aspx?MID=17555325&aid=17555326&pIType=read_it
- [6] MUSE. (n. d.). Operations Improvement Toolbox. Retrieved from MUSE: https://class.ctuonline.edu/_layouts/MUSEViewer/Asset.aspx?MID=17555246&aid=17555271
- [7] STACKIFY. (2018). What Is SDLC? Understand the Software Development Life Cycle. Retrieved from What Is SDLC? Understand the Software Development Life Cycle: <https://stackify.com/what-is-sdlc/>
- [8] Stewart, D. (2012). The Real Problem with ECM. Retrieved from The Real Problem with ECM: <https://blogs.gartner.com/darin-stewart/2012/05/04/the-real-problem-with-ecm-hint-it-isnt-the-platform/>
- [9] Thomas, G. (n. d.). Defining Data Governance. Retrieved from Defining Data Governance: <http://www.datagovernance.com/defining-data-governance/>
- [10] Thomas, G. (n. d.). The DGI Data Governance Framework. Retrieved from The DGI Data Governance Framework: http://www.datagovernance.com/wp-content/uploads/2014/11/dgi_framework.Pdf
- [11] Wiessner, A. (2017). The Benefits of Enterprise Content Management and Software as a Service. Retrieved from The Benefits of Enterprise Content Management and Software as a Service: <https://identi.net/benefits-moving-enterprise-content-management-cloud/>