Enterprise Data Management (EDM) helps Communication Service Providers (CSPs) address the challenges caused by convergence of technologies and frequent mergers and acquisitions. It provides a single view of the truth, unique reference data and a unified data quality framework to integrate, validate and migrate data. In addition, it enables continuous monitoring of the quality of data and establishes standards across the enterprise data lifecycle. Our white paper discusses the need for EDM in the telecommunications sector, the benefits and challenges in implementing EDM, and the components of an effective enterprise data management solution.
Executive Summary

Mergers and Acquisitions (M&As), convergence of technologies and rapid changes in global regulations are revolutionizing the telecommunications industry. To address the dynamic market conditions as well as the competition from established companies and new entrants, Communications Service Providers (CSPs) must constantly provide new features and opportunities to customers. In addition, they must foresee market changes and act decisively by capitalizing on their primary asset – operational and analytical data – for precise metrics and predictions. However, CSPs must implement and govern enterprise data effectively to meet their business intelligence needs.

Enterprise Data Management

Enterprise Data Management (EDM) helps CSPs manage heterogeneous data sources, validate the quality of data, devise a common data model by integrating information, build analytical and presentation layers, and manage end-to-end metadata in the analytical and presentation layers. EDM also guides and governs the data architecture, while managing data assets. A typical EDM architecture diagram is depicted in Figure 1.

Benefits of EDM

EDM ensures consistency of information with a ‘single version of truth’ by providing reference data requirements on an integrated data platform. It supports operations and enhances decision-making capabilities by helping CSPs migrate from disparate data silos to an integrated, enterprise-wide data environment.
EDM delivers several benefits -

- Provides a single, accurate view of end-to-end enterprise data
- Consolidates, profiles and integrates data from disparate systems, thereby enhancing Data Quality (DQ) and eliminating exceptions. It increases the trust factor of data.
- ‘Insulates’ systems and processes from change, enabling rationalization of legacy systems and facilitating mergers and acquisitions
- Improves the accuracy of decisions
- Enables on-demand extraction of ad hoc operational reports
- Facilitates rapid analysis of customer retention, customer satisfaction, customer lifetime value, and cross-sell and up-sell data
- Enables risk-benefit analysis of business opportunities
- Enables reuse of the EDM framework in new regions, markets and product categories

**Dimensions of Enterprise Data Management**

![Image of enterprise data management framework]

The components of an enterprise data management framework are shown in Figure 2, and are discussed in subsequent sections. Enterprise data can be managed along various angles/ layers to meet the goals of EDM:

- Data stewardship
- Data governance
- Data standards
  - Data definitions and taxonomies
  - Technology standards
  - Data retention
- Data quality management
- Data architecture
  - Data integration
  - Data migration
  - Master/ reference data management
- Metadata management
- Data warehousing
- Data portal
- Data security
Data Stewardship

EDM requires the owners of source data to manage data assets effectively. A data steward’s responsibilities include managing data standards, formats and trust factor, and establishing and enforcing data standards. Data quality management provides quality analysis reports that enable stewards to improve the quality of data, reduce data redundancy and improve data management capabilities across the enterprise.

Source data owners provide the overall governance and oversight to the data management activities within an enterprise, while data stewards -

- Define business terms and establish rules for quality and exception validation Identify business-critical attributes and ensure data validity
- Prioritize and ensure data quality of exceptions raised by the quality management team
- Coordinate between business, application and IT teams to enhance the trustworthiness of data
- Address target and downstream data requirements
- Create awareness about data management principles, data security and retention policy, and best practices across the enterprise

Data Governance

A data governance committee must be instituted to develop the principles of managing data-related processes and enforcing them across the enterprise. It must be responsible for nominating teams for programs such as data stewardship, data quality management, data modeling, data integration, data migration, data warehousing, master data management, data architecture, data security, and metadata management. A data governance framework that focuses on people, process and technology ensures accessibility, availability, quality, consistency, security, and audit-readiness of data.

Typically, senior managers form the steering committee, while the operational team of the data governance committee includes business users, data architects and data stewards. Data governance plans are framed by a working group and approved by the steering committee. The committee must ensure that data assets are managed effectively across the enterprise.

Key roles and responsibilities of the data governance committee:

- Steering committee – Articulate the vision and arrange funds for the data governance initiative
- Source data owners – Prioritize and execute data management
- Data stewards – Address issues in data quality and standards such as merger or deletion of data, data enrichment, etc. Data stewards must ideally be from the business side.
- Data architects – Help data stewards access, integrate and manipulate data with their technical expertise
Data Standards

Data standards are framed by the data governance steering committee to ensure that all data elements of an enterprise comply with standard terms, definitions and values. The working group, in turn, ensures that all associated parties agree with the data values and adhere to the standards. It helps the data stewards frame validation rules to filter business exceptional data and improve its trustworthiness. DQ issues identified by validation against standards must be closely monitored and tracked by data stewards until resolution.

By ensuring that each and every data element of the enterprise data model adheres to the rules and definitions of the SID model, CSPs can meet the data requirements of emerging technologies such as wireline, wireless, cable, and IPTV. In addition, they can avoid redundant data and ensure consistency of data.

Data elements can be classified based on their logical and physical properties. Logical property refers to the definition, origin and data type. Physical properties of the data element include data length, validation rules, and how data is stored, presented to end users and labeled. For example, elements such as the date and currency fields have specific display formats, standards, regional settings, and validations.

Data Quality Management

Data Quality Management (DQM) is a key enterprise data management process that addresses issues in the quality of data, and identifies exceptions in data elements that can be classified into industry standard quality dimensions as shown in Figure 4. It requires strategic data profiling and data quality software to be a part of the data management process.

Figure 4: Classification of exceptional data elements

---

1 The Shared Information/Data Model (SID), framed by the TM Forum, provides a single set of terms for business entities and attributes across the telecommunications industry. SID enables business users to use the same terms to describe business objects, practices and relationships across the enterprise.
The quality of data may be affected at various stages of the data element lifecycle such as data entry, data transformation, conversion from operational to analytical data, master data and reference data transformation, and migration of data from legacy systems. Some practical examples of anomalies in data quality:

- **Accuracy**
  - Mary Pierce, 1408 North Any Street, Germantown, MD, 418-734-1576
  - *Is this a valid postal address?*

- **Consistency**
  - Alan Smith, 4200, Weston Park, Cary, NC
  - *Is there a suffix to the street?*
  - *Is there a postal code?*

- **Completeness**
  - Street => St, Str
  - Account => Acct, Accnt, A/c
  - *Are they business terms with variations in value?*

- **Correctness**
  - Mary Brown, 1408 North Any street, Germantown, MD, 9AB-786
  - *Is this a valid zip code format?*

- **Uniqueness**
  - 'First Merit Bank' and '1st Merit Bank'
  - *Are they the same or different*

- **Relevance**
  - Robert Smith, 123 Peach Tree lane, Pleasanton CA
  - Julia Smith, 123 Peach lane, CA
  - *Are the two from the same household?*

- **Timeliness**
  - Activation date and time: 12 May, 2009, 12:00 AM, posted into the system during day-end batch processing. However, the service was activated during business hours of the previous day.

- **Interpretability**
  - First name attribute contains: Tai-Tai Cheung Mei Lee Wang
  - *Is this a personal name or company name?*
  - *Is the gender male or female?*

Data management tools can be used for profiling and standardizing data, matching and merging data, monitoring quality, and tracking and addressing issues in data quality.

**Challenges in Data Quality Management**

- **Ownership of quality issues**
  - Continuous monitoring of quality by the DQM process enables real-time reporting of issues. However, it is seen that units/ departments rarely take ownership to rectify them or approve automatic cleansing. A data governance committee is needed to enforce cross-functional collaboration and sensitize the organization about the importance of resolving DQ issues. It requires huge manpower, and the Return on Investment on data quality fixes is relatively low.

- **Identifying master/ reference data**
  - The lack of attribute-level standards for data quality, M&As without a data integration strategy and federated source data stores result in multiple versions of data, the lack of a ‘golden’ record and incomplete master attributes.

- **Metadata-driven DQM**
  - A DQM process driven by metadata management is recommended for CSPs with frequent Merger and Acquisitions, multiple source systems or strategic plans to integrate data from multiple technologies. The cleansing maps in the integrated process and its in-built mechanism to alert and track issues help refine DQ, exception and notification rules and resolution techniques. The DQ system notifies exceptions to quality standards and violation of data quality rules to data stewards, who can rectify the issues with the help of data architects. New DQ issues must be updated in the rule repository. Quality analysis reports contain the results of DQ measurements in a pre-determined format. The senior management normally requires reports to be presented with color codes akin to the traffic signal – usable data (green), partly usable (yellow) and not usable (red).
• **Data quality-driven cleansing**

The DQ process meets business requirements by cleansing and standardizing the database through continuous monitoring. The data governance committee and data stewards frame the rules for cleansing, exception management and standardization of data. Data quality scorecards help data stewards monitor and track DQ issues, and refine data cleansing rules. Cleansing occurs at various instances of the data management lifecycle:

• Front-office (real-time) cleansing
• Back-office (batch cleansing)
• Cross-office (cleansing during data transfer between businesses)

• **Data collaboration between cross-functions**

The DQM process is burdened with ambiguity and risk when an attribute has multiple owners or a change in an attribute influences other attributes and downstream applications. It can be avoided by storing details of the creation of data in the taxonomy of data items. The information to be stored include author, date and time of creation, application details, domain values, derivation details, derivation logic (if required), dependant fields, data item hierarchy, and propagation of changes.

As a DQM best practice, CSPs must start with limited data profiling and data cleansing activities, and incrementally develop a robust and scalable data quality management platform for a cross-organizational, 360-degree view of business.
Data Architecture

Data architecture is a fundamental aspect of enterprise data management. It is a multi-layered set of models that defines the enterprise data strategy and management policy on data collection, and identifies the need for improvement in business decisions.

The enterprise data strategy comprises strategic initiatives such as data integration, legacy data migration, legacy system rationalization, master data management, metadata management, and business intelligence and reporting.

CSPs can integrate independent applications into an ‘enterprise data model’ to address challenges such as dynamic business processes, convergence of technology, changes in regulations, and increased competition. In addition, they can view and monitor enterprise performance by integrating and migrating data into the enterprise model, enabling rationalization of legacy systems.

Data architects design the enterprise data model, automate data capture and validation, and implement an audit trail mechanism for business-critical data. They are also responsible for:

- Entity relationship diagrams that depict the relationship of business entities across subject areas
- Data flow diagrams that display data flow between enterprise applications and databases
- Identifying data stewards across subject areas, associated business units, business processes of each business function, and enterprise applications
- Setting standards and best practices for naming conventions to define data elements in the enterprise data model
- Designing and managing a metadata repository across the enterprise. The repository stores attribute name, type, length, owner of the data element, business unit, valid values, last updated user, last updated date and time, etc.

Data profiling; data modeling; Extract, Transform and Load (ETL); and reporting tools exchange metadata with the enterprise metadata repository.
The aim of the enterprise data strategy must be to provide cleansed, consistent, integrated, and well-managed reference or master data. Since master data is a core component, any issue with it will affect the entire enterprise, exposing risks due to data inconsistency.

**Data Security**

CSPs need a robust strategy to ensure security of sensitive, business-critical data such as customer profile, payment details, contact, contracts, subscription, and product details. The data governance team must frame rules and regulations to –

- Manage the change control board that authorizes changes to the data structure for sensitive data. Frequent changes lead to an unstable business and multiple versions of the business entities.
- Enhance the confidentiality and availability of data in hard and soft copy
- Protect data from unauthorized access, modification and destruction
- Prevent improper disclosure of data
- Avoid security breach of information and related loss to business, legal implications, etc.

The data security strategy assigns stewards for all data sources, and authorizes them to grant access rights, maintain the list of authorized users and ensure accuracy of data. Stewards must also ensure that data is not duplicated in any format unless there is a business process requirement and copies are controlled across the enterprise.
Challenges in Enterprise Data Management

To manage terabytes of enterprise data in a complex landscape of legacy applications, CSPs require an enterprise data model that is compliant with the SID framework. They also need well-defined strategies for data integration, modernization of applications, migration of data from legacy applications, rationalization of legacy systems, and presenting data in the enterprise portal. It involves the following challenges:

- Duplication of data due to heterogeneous applications with independent data systems
- Data impurity due to the lack of data quality principles across the enterprise
- Non-standard data and exceptions in data range, type and length due to the absence of a metadata platform
- Multiple owners maintain data pertaining to different technologies in various data structures
- Lack of a data governance committee to enforce data quality principles and quality standards, and create awareness among source system owners about the importance of data as corporate assets
- Multiple versions of operational reports with different sub-functions and duplication of data in different formats across upstream/downstream data systems due to uncontrolled data distribution, resulting in legal and regulatory non-compliance

Conclusion

Enterprise data management helps CSPs improve the quality of data, prevent revenue leakage and roll out methodologies for data governance, metadata management, master data management, data architecture, and data security. It also enables informed decision making through customer behavior analysis, market behavior analysis, competitor analysis, single view of enterprise data, and converged billing. However, the success of an EDM implementation depends on effective collaboration between business sub-functional heads, data architects, the CIO and CEO.

About the Author

Sivaprakasam S.R. is a Principal Technology Architect and mentors the database and business intelligence track at Infosys' Communication, Media and Entertainment business unit. His interests include enterprise data modeling, enterprise data integration, enterprise data warehousing, enterprise data quality management, and semantic data integration. He can be reached at sivaprakasam_s@infosys.com