ISSN: 2583-9888
DOI: doi.org/10.51219/JAIMLD/ravi-kumar-vallemoni/656

URF PUBLISHERS
connect with research world

Journal of Artificial Intelligence, Machine Learning and Data Science

https://urfpublishers.com/journal/artificial-intelligence

Vol: 4 & Iss: 1

Research Article

Post-Merger Canonical Enterprise Data Model (EDM) for Banking

Ravi Kumar Vallemoni*

Citation: Vallemoni RK. Post-Merger Canonical Enterprise Data Model (EDM) for Banking. *J Artif Intell Mach Learn & Data Sci 2021* 4(1), 3250-3258. DOI: doi.org/10.51219/JAIMLD/ravi-kumar-vallemoni/656

Received: 02 March, 2021; **Accepted:** 18 March, 2021; **Published:** 20 March, 2021

***Corresponding author:** Ravi Kumar Vallemoni, Senior Data Architect - Bank Of America, USA

ABSTRACT

The banking sector has been facing serious data management challenges as a result of post-merger integration (PMI) where there is a need to harmonize different heterogeneous customer, account and transaction data models which have many origins of different legacy institutions. Separated data architecture, mixed up semantics and varying governance activities often lead to broken reconcilement, slow financial close process, greater operational risk and regulatory compliance gaps. The paper suggests a Post-Merger Canonical Enterprise Data Model (EDM) of banking that is based on an Authorised Data Store (ADS) pattern of a pattern that manages, certifies and attests a golden source of enterprise data. The presented framework proposes certified data snapshots and formal attestation processes and service-level agreements (SLAs) that are specific to the downstream consumers in the finance, risk, liquidity and regulatory reporting functions. The study presents the known data warehousing concepts, master data management (MDM) and financial data governance techniques as a single canonical paradigm that normalises the customer, account and transaction entities across the merged institutions. Semantic consistency and auditability is stipulated with the definition of reconciliation rules, data contracts, lineage controls and golden-source governance mechanisms. Findings indicate that the implementation of the suggested ADS-centric canonical EDM is associated with a significant decrease in downstream breakages along with the improvement in close timeliness as well as the reduction in manual rework in all financial reporting processes. The article has added a practical reference set of architecture and governance blueprint relating to banks that are merging, helping regulate them, operational performance and scalable enterprise analytics.

Keywords: Enterprise Data Model, Post-Merger Integration, Banking Data Architecture, Canonical Data Model, Authorised Data Store, Data Governance, Financial Reporting.

1. Introduction

1.1. Background

The acquisition and merging have been a strategic instrument where banks aim to grow at a faster rate, expand their market and product and geographical diversification[1-3]. Although strategic in terms of intent, the post-merger integration phase has always been the most intricate and failure-nine stage of the consolidation process with data integration difficulties mentioned among the most critical inhibitors of the timely value delivery. This was the case in the years before March 2021 when most banks had very heterogeneous technology environments consisting of decades of organic growth, systems improvement driven by regulations and acquisitions. Regional and business line differences in core banking platforms were enormous and frequently represented local regulatory demands and vendor antiquities. The customer master data was generally spread out to multiple repositories that are all optimized with various operational or compliance requirements and transaction processing platforms were also often customized and product

or channel specific. The fragmentation led to inconsistent definition of data and duplication of records and transparency within the enterprise. Consequently, the effort to unify data became mostly reactive, expensive and time consuming after mergers, postponing integration synergies and augmenting the operational and regulatory risk. It is these enduring problems that lead to seeking more formalized, controlled and scaled down data integration strategies that can better help in post-merger banking settings.

1.2. Importance of post-merger canonical enterprise data model

A Post-Merger Canonical Enterprise Data Model (EDM) is an important component in balancing and rationalizing the data landscape after a banking merger or otherwise acquisition. Without a shared semantic base, merged institutions tend to be plagued by data definition inconsistency, redundant reconciliation business logic and fragmented reporting results. Adopting a canonical EDM gives a unified language of data across the legacy systems that allow it to be readily interpreted, easily integrated and governed scalably **(Figure 1)**.

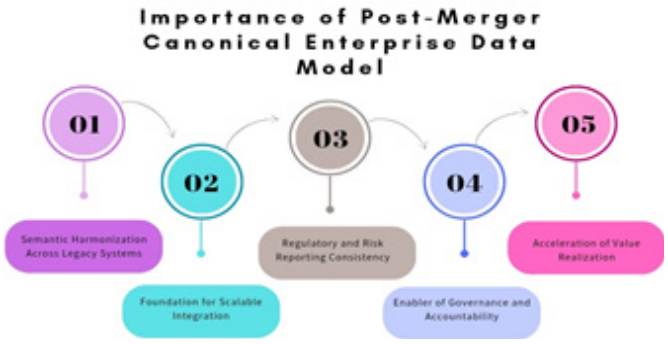

Figure 1: Importance of Post-Merger Canonical Enterprise Data Model.

- Semantic Harmonization Across Legacy Systems: Another most obvious advantage of a post-merger canonical EDM is that semantics become harmonized across heterogeneous legacy platforms. The standardizations of what is meant by core entities, including customers, accounts and transactions, mean that the model results in ambiguity being removed as well as misinterpretations between systems within the same entity that have been inherited by other institutions. Such semantic agreement is critical in correct data aggregation, comparability and consolidation on post-merger settings.
- Foundation for Scalable Integration: Canonical EDM is a consistent integration layer that separates the complexity of a system on the upstream side of the integration point and downstream customers. Rather than creating several point-to-point mappings among the old systems, the data is mapped once into the canonical model and reused in analytical, financial and regulatory applications. This strategy decreases the complexity of integration, lowers the maintenance price and hastens the rate of rationalization of post-merger systems.
- Regulatory and Risk Reporting Consistency: The increased regulation of post-merger banks makes the provision of consistent, traceable and auditable data of business operation in the enterprise inevitable. A canonical EDM offers the skeletal support required to enable the aggregation of risks, the computation of capital and regulatory disclosures. Conforming the data definition to regional concepts facilitates consistency in reporting and lowers the potential of regulatory discoveries involving the quality and origin of data.
- Enabler of Governance and Accountability: In addition to technical integration, a post-merger canonical EDM can permit increased data governance by developing a clear definition of data domains, ownership and data steward responsibilities. It advocates the use of certification, attestation and service-level agreement by offering an authoritative source of what data is and its use. This management ability turns information into an unorganized operational waste to an organized business resource.
- Acceleration of Value Realization: With an early fix of semantic inconsistencies and offering a reuseable integration base, a canonical EDM speeds up the achievement of merger synergies. Consistent and trusted data can be delivered to business and analytics teams more quickly, allowing them to make quicker decisions, gain better customer insights and have more reliable financial reporting. Consequently, the canonical EDM emerges as a strategic facilitator of the success of post-mergers instead of a technical object.

1.3. Banking: Unifying customer, account and transaction models across institutions

The most complicated and subject to ramification of post-merger integration in the banking industry is the challenge of customer, account, transaction models integration among merging institutions[4,5]. These core entities are traditionally defined, hierarchized and with business rules within each institution and are usually influenced by legacy systems, regulatory interpretations and product strategies. Consequently, what appear to be precisely the same concepts, a worker who is described as a customer, an account or a transaction, can vary in their meaning, form and level of detail across firms. In the absence of a common model, these mismatches do spread into downstream processes, resulting in inconsistency in reporting, duplicated logic of reconciliation and less trust in enterprise data. The major challenge in unifying customer data is associated with the differences in identification schemes, relationship models and regulatory classifications. A single institution might capture a customer at the personal level and another at the legal or household level which makes it hard to report on customers and assess risks as a single entity. Likewise, account models may vary in how they represent product structures, ownership and ledger representations. Merging of institutions with disparate accounting practice or system designs makes it especially challenging to reconcile the operations accounts with accounting constructs. The complexity is further intensified in transaction models because the differences in the sequence of events or posting logic, settlement timing and historical retention policy may have a significant impact on balances, profitability analysis and regulatory metrics. The canonical model serves to overcome these challenges by modelling customer, account and transaction concepts as more standardized and enterprise-wide representations that do not depend on the source-system idiosyncrasies. This abstraction allows the same aggregation, traceability as well as interpretation throughout the combined organization. The unified model is useful in risk aggregation, regulatory reporting and enterprise analytics by using it to act as a common semantic base. More to the point, it helps banks to

go beyond temporary integration solutions to a long-term data architecture that facilitates long-lasting scalability, governance and relationship of confidence in post-merger settings.

2. Literature Survey

2.1. Enterprise data models in banking

Enterprise Data Models (EDMs) have found a lot of use in the banking sector to have a standard and coherent representation of the fundamental business notions within the heterogeneous systems and activities. The initial industry-oriented EDMs[6-8], which were frequently being propagated by consulting companies and standards organizations, focused on logical abstraction and semantic integrity of the underlying entities like customers, accounts, products and transactions. These models were helpful in increasing interoperability as well as minimizing the ambiguity in data interpretation across business lines. Nevertheless, the literature shows that the majority of EDM efforts were still conceptual in nature relying on design, as opposed to run-time enforcement. Such governance mechanisms as the accountability of ownership, certification of usage of data and operational controls, to a great extent, were not established and restricted the usefulness of EDMs in real-life banking operation. Consequently, although EDMs led to better mutual understanding, the data were not always consumed similarly and with confidence across enterprise platforms.

2.2. Post-merger data integration challenges

The literature has popularly attributed post-merger integration (PMI) as one of the most stressful areas of enterprise data architectures and mostly in the banking industry. The literature always points to semantic divergence of merging institutions where the same data items in meaning, structure or granularity are different. Data duplication and survivorship issues also make integration complicated because two or more systems will claim control over one customer or account record. Moreover, historical data retention policies vary across time make it difficult to reconcile legacy data sets, particularly when comparability across time is needed by regulation reporting requirements. The technical difficulties are compounded by increased regulatory scrutiny of the post-merger era businesses because banks need to quickly show their data accuracy, lineage and consolidation talents. The literature highlights that the long old methods of integration, which tend to be motivated by the short-term operational requirements struggle to provide sustainable and compliant data base in the after-merger scenarios.

2.3. Golden source and master data management

The idea of having a golden source or having a centralized hub of the Master Data Management (MDM) evolved due to the continual inaccurate data and a disjointed data on a banking system. According to academic and practical literature, golden source architecture are authoritative depositories aimed to deliver a single and trusted variant of very important reference and master data. Although these implementations enhanced consistency in structure, the majority were labour-intensive with vast use of batch-oriented type of synchronization and periodic reconciliation. Notably, mastered data were frequently reinterpreted or overridden by downstream systems by which trust in the central repository was undermined. It is mentioned in the literature many times that there are no formal attestation, certification or other accountability structures that would hold consuming systems accountable to recognizing and following the golden source. As a result, even tremendous investments did not help MDM initiatives to end data trust issues properly and became another layer in an already sophisticated data landscape.

2.4. Regulatory drivers

Regulations have also been instrumental in the current enterprise data practices in the banking sector, especially after the global financial crisis. The principles of data accuracy, completeness, timeliness and traceability in risk reporting and aggregation were highlighted in regulations released before 2021, including the most significant one, which is BCBS 239. According to academic research, as much as these regulations spelt out what was intended to be realized, there was a lack of direction as to what data architectures would be needed to make it a reality. Consequently, banks mean compliance differently, tending to have piecemeal solutions to certain regulatory reports instead of tackling underlying data governance loopholes. This regulatory mechanism yielded results of local success in compliance, but it did not provide incentive to accelerate comprehensive (enterprise-wide) data operating model, leaving gaps in the systems unsolved.

2.5. Research gap

Although thoroughly studied business data model, integration after a merger and master data management have been studied and examined extensively, the literature provides a severe lack of insight into integrated practices. The current literature scarcely hypothesizes models that concomitantly integrate a canonical enterprise data model and operationally entitled information stores that are personally defined to consume. In addition to that, the so-called formal data attestation, i.e. certified data sources and service-level agreements (SLAs) and responsibility of quality and availability is mostly missing. It is this gap that indicates the need to have a single framework that goes beyond conceptual standardization and into enforceable data trust that will bridge the gap between data modelling, governance and operational implementation in regulated banking systems.

3. Methodology

3.1. Canonical enterprise data model design

Canonical Enterprise Data Model (EDM) is a framework model aimed to give a stable and unified semantic platform that can support enterprise-wide data interoperability, regulatory compliance and analytical consistency[9-11]. It represents concepts in core banking in standardized areas that are not related to application-dependent implementation so that the interpretation can be uniform and evolution can be guided across systems. The model is structured based on three core domains of Customer, Account and Transaction representing the main dimensions of the financial activity and the regulatory reporting **(Figure 2)**.

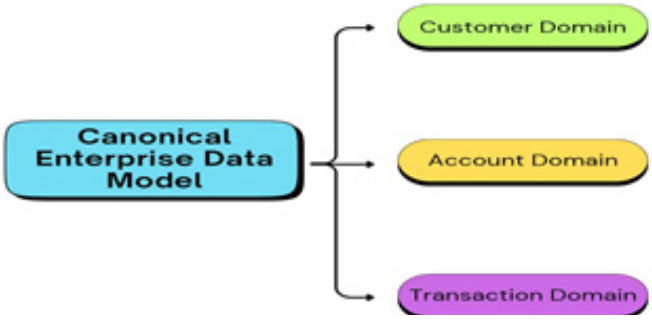

Figure 2: Canonical Enterprise Data Model Design.

- Customer Domain: The Customer Domain creates a standardized image of the parties and relations within the enterprise. It separates individuals and legal entities and promotes the higher-order group structures like households and the hierarchies of organization. The entities of relationship document relationships, including beneficial interest, ownership, control and authorization. This domain facilitates proper identification of customers, know-your-customer (KYC) and anti-money laundering (AML), as well as provides a uniform customer centric reporting and risk aggregation across business lines..
- Account Domain: Account Domain is a model that represents the financial instruments and forms in which the value is stored, tracked and reported. It distinguishes between product accounts, representing customer facing financial arrangement and ledger accounts, which reflect the accounting expression consistent with the general ledger arrangements. Balance entities present time perspectives of account balances whereas the ownership constructs determine an accountable entity. This separation promotes multi-product banking operation, provides good reconcilment of operational and financial system and ensures uniformity in financial reporting and calculation of capitals.
- Transaction Domain: The Transaction Domain signifies the life cycle of financial transactions, starting with the point of initiation to the point of settlement. Financial events reflect this economic purpose of a transaction, whereas postings show the accounting effect across all accounts in which they are relevant. Settlement entities identify value transfer achievement, which are timings, counterparties and clearance mechanisms. This field is beneficial in terms of traceability and auditability, as it ensures that business events and accounting outcomes have a clear lineage, which is required to comply with the regulations, dispute and calculate risks.

3.2. Authorized data store (ADS) pattern

Authorized Data Store (ADS) pattern is a managed and regulated layer of data consumption, which is the only validated source of downstream analytical, reporting and regulatory use cases within the enterprise. As opposed to traditional data warehouses or data lakes[12-14], where people tend to mix data of several different sources with varying data qualification and authority, the ADS is specifically defined as the system of analytical record. Its main aim is to operationalize trust, which will see to it that all information that will be used to make decisions and submit information to the regulators are of a formally authorized, traceable and quality assured repository. The ADS reduces the ambiguity on data provenance by defining one point of certification and removes multiple interpretations of critical financial and risk data. Another characteristic of the ADS is that it uses immutable certified snapshots. When data has been validated, reconciled and certified at a specified period in time or cut-off analysis point, it is not overwritten. Any further amendments or changes are logged as new versions and not in place which maintains accuracy and auditability of the past. It is this immutability that allows regulators and internal stakeholders to replicate previous reports as they were relayed, which can solve long-standing issues in restatement and audit defense. ADS is dynamically time sensitive and the effective dating is encoded throughout all the core entities. The design will enable the store to reflect up-to-date states of data, both historical states as well as useful in as-of reporting, trend analysis and regulatory lookback. Effective start and end dates will guarantee that modifications in customer characteristic, customer account structure or transactional conditions are contextualized in a suitable approach to time which is central to risk summation, capital calculations and compliance reporting. The ADS also provides embedded metrics of reconciliation which makes a big difference between the ADS and standard data repositories. These measures do not only measure completeness, accuracy and alignment of the source systems with the canonical form and certified outputs. The results of the emerging reconstruction are stored with the data and this introduces visibility of the data quality, which can be formally attested using the service-level agreements (SLAs). Together, these properties make the ADS a base of architectural pattern to have scalable and regulator-ready data trust.

3.3. Certified snapshot mechanism

The certified snapshot mechanism is a formalized mechanism of generating trusted, point-in-time enterprise data representations that are appropriate to regulatory reporting, aggregating risk and enterprise analytics. Snapshots are determined at stipulated cut-off points e.g. end-of-day (T+0) or post-reconciliation windows (T+1) to reconcile with operational and regulatory cycles. A snapshot will capture the entire and unchangeable state of the Authorized Data Store at a given point of time and maintain consistency among all lower-level consumers. The mechanism also helps to remove variances in changes in timings of the snapshot by standardizing timing of the snapshot, allowing repeatable auditable reporting cycles with asynchronous data extraction. The model of certification is in a form of quantitative scoring, which considers its completeness as well as quality of data. The Certification Score is then obtained in an empirical manner, dividing the number of records that are actually validated by the overall number of records in the snapshot then multiplying the ratio with an aggregate Data Quality Index. The validated records are those records that have successfully gone through all the stipulated business rules, reconciliation and referential integrity checks. The overall number of records corresponds to the anticipated number of people in accordance with the source systems of the upstream and the definitions of the canonical models. Data Quality Index is used to combine several dimensions of quality (accurate, consistent, timely and compliant with Canonical Enterprise Data Model) into a single score that is weighted. This mixed solution will make sure that a snapshot is not certified based on the volume or completeness but you are guaranteed the quality of the data contained in it. A snapshot that has high coverage of records but of low quality or vice versa will not qualify as being certified. Those snapshots that surpass some predetermined certification thresholds are officially published to be consumed by regulatory, financial and analytical systems. Snapshots that do not satisfy the threshold are put on quarantine to remediate with diagnostic measures being provided to data owners and stewards. The snapshot mechanism provides a mechanism to operationalize data trust and introduces the concept of accountability into the lifecycle of the data by implementing objective, metric-driven certification. It offers apparent adherence, facilitates auditing defensibility and forms a repetitive basis of service-level agreements regulating enterprise data consumption.

3.4. Attestation workflow

Attestation workflow creates a formal and auditable process where data is reviewed, confirmed and finally certified to be used in the enterprise and regulations[15-17]. It creates transparency of accountability because it achieves a level of explicit accountability by also engaging the technical and business stakeholders in certifying data fitness, transferring data trust, which is an implicit assumption, to operational control. The work process is implemented in a sequential order and each stage has quantifiable results and evidence that can be traced **(Figure 3)**.

Figure 3: Attestation Workflow.

- **Data ingestion validation:** Data ingestion verification is a technique that guarantees that the incoming data into the Authorized Data Store is full, in time and free of technical errors. This phase confirms that there was successful data transfer between the upstream source systems, record counts, file integrity, schema compliance as well as timeliness during ingestion is checked against pre-established expectations. This phase of failures means that extraction or transportation is a problem that must not be processed downstream until fixed and that uncomplete or corrupted data should not be forwarded into certified snapshots.
- **Structural reconciliation:** Structural reconciliation authenticates the correspondence existing between the information consumed and the Canonical Enterprise Data Model. It ensures that compulsory attributes are filled in, data types and formats are appropriate and referential integrity maintained between related entities like customer, accounts and transactions. This step isolates structural anomalies, e.g. orphan records or invalid relationships, that may compromise on accuracy of aggregation or reporting when not corrected.
- **Semantic reconciliation:** Semantic reconciliation determines whether data is in agreement with the agreed enterprise definitions and business rules. These involve authenticating code values, classification logic, calculation procedures as well as contextual interpretations across areas. Indicatively, it also makes the balance of accounts, type of transactions and type of customers to have consistently been derived over the source systems. There is a strong need to provide semantic reconciliation in order to eradicate interpretive irregularities that tend to remain despite structural congruity.
- **Business sign-off:** Business sign-off is the last attestation process and the certified snapshot must be approved by allowing specific data owners (or responsible business executives) to approve the certified snapshot. The approval assures the data of conforming to agreed quality criteria, regulatory requirements and business utility standards. The accountable data certification is registered as auditable evidence, with data certification participants linked to responsible roles and service level agreements. This will guarantee that data quality is owned by the organization and also build trust with the consumption downstream.

3.5. Data contracts and SLAs

Data contracts and service-level agreements (SLAs) capture the relationship between data consumers and data producers by providing explicit and enforceable expectations regarding how the data will be delivered, how it will meet quality expectations and how the data will be accountable. Instead of data exchange being a casual or best-effort event, this framework makes data an enterprise asset that has quantifiable responsibilities. Contracts are harmonized with Canonical Enterprise Data Model and Authorized Data Store, which provide consistency, reliability and compliance with regulations throughout the data lifecycle **(Figure 4)**.

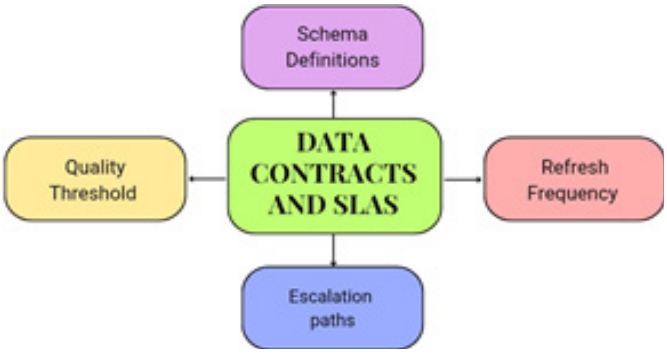

Figure 4: Data Contracts and SLAs.

- **Schema definitions:** Schema definitions define both the structural and semantic model of data between producers and consumers. Their definitions specify entity structure, attribute names, data types, allowable values and relationship constraints following the structure as presented in the canonical model. Data contracts have the effect of eliminating breaking changes in a schema and also making data comprehensible by consumers in a reliable way since data contracts explicitly version and manage schemas. This transparency is necessary to ensure stability of downstream analytical and regulatory systems.
- **Refresh frequency:** Refresh frequency determines the frequency of data delivery/updates to the Authorized Data Store like intraday, end-of-day or events. This requirement guarantees the availability of data in line with business and regulatory schedules, such as reporting deadlines and risks aggregation periods. Well defined refresh periods minimize uncertainty, stale data use; and allow consumers to design processes, which rely on predictable data delivery.
- **Quality thresholds:** The quality thresholds set the lowest level of completeness, accuracy, consistency and timeliness of data. These thresholds are directly associated with certification measures and reconciliation data and outline objective standards by which data have to satisfy, before it is made available to be consumed. When quality expectations are included in SLAs organizations change the data quality

to the problem-solving tool and advance to the proactive tool of performance management.

- **Escalation paths:** Escalation tracks establish the process and the responsible position of data problem resolution in case the contractual mandatory is not fulfilled. They define Notification schedule, data owners who are in charge of such remediation or exception management. This guarantees the systemic handling of data incidents which is transparent and accountable which adds strength to audit and regulatory scrutiny.

3.6. Reconciliation rules

Reconciliation rules are used to offer the systematic controls to help maintain consistency, completeness and accuracy of data transfer among the source systems into the Authorized Data Store and certified snapshots. The rules run at different data model levels and allow identifying the discrepancies as early as possible and quantitatively demonstrate the data integrity. The framework facilitates traceability and defensibility requirements in the regulatory aspects of the data lifecycle, through the introduction of reconciliation into the data lifecycle.

- **Record count reconciliation:** Record count reconciliation confirms that the number of records put into the Authorized Data Store are within the expected amounts of records that were put into the record by the source systems. It is an algorithm that matches the source extracts, ingestion logs and canonical representations to determine what records are missing, duplicate or unexpected. Any variance outside the designated tolerant limits initiates the research process before the certification of any data, so that ultimate consumers do not work with incomplete or exaggerated data.
- **Balance aggregation reconciliation:** Balance aggregation reconciliation provides consistency in financial totals in both higher and lower aggregations and boundaries of systems. Balances at account level in the Authorized Data Store are compared against those in the source system and translated to upper level sums, such as product, portfolio or general ledger balances. This control assures that financial accuracy is maintained in the transformations, currency conversions and other adjustments and this is important to regulatory reporting, capital adequacy and other financial disclosures.
- **Cross-domain referential integrity:** Cross-domain referential integrity reconciliation ensures that the relationship between the core domains, including customers, accounts and transactions, are entire and valid. This involves ensuring that all the accounts are associated with a valid customer and that all transactions are in reference to an existing account. Orphaned references or invalid references during detection of these references avoid analytical distortions and forms good risk aggregation as well as customer-friendly reporting across domains.

3.7. Governance and golden source management

Management and governance assistance offers the foundation of the organization that is needed to support the Canonical Enterprise Data Model and the Authorized Data Store in the long term[18-20]. It has a centralized data governance council to see that data requirements, requirements and data sources remain strategic hence to business strategy, business reality and business regulation. This council is the decision-making organ that enforces consistency and limited evolution, to ensure that enterprise data assets are not fragmented **(Figure 5)**.

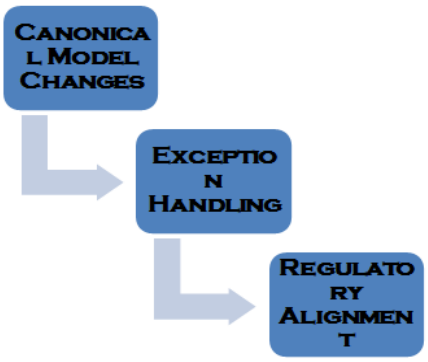

Figure 5: Governance and Golden Source Management.

- **Canonical model changes:** All modifications to the Canonical Enterprise Data Model such as the addition of new entities, attributes or relationships or changes to the definition of existing ones are controlled by the governance council. Before approving change requests, they are assessed in the following areas: business justification, impact on downstream and regulatory implications. The managed change process maintains the stability of the models and allows evolution to adapt to new products, regulations or analysis needs.
- **Exception handling:** Exception handling governance determines the way of management of exceptions to normal rules of data, quality and certification standards. The council has policies on accepting temporary and permanent exceptions, records the rationale, scope and remediation plans. Through the formalization of exception management organization does not perform workarounds but creates transparency and audits in the use of data.
- **Regulatory alignment:** Regulatory harmony is used to maintain the golden source and related practices of governance as being in accordance with emerging supervisory requirements. The council deciphers regulatory requirements and converts them into data standards, reconciliation rules and certification requirements. Such proactive alignment helps to minimize regulatory risk, facilitates uniform reporting and allows responding promptly to regulatory changes or inquiries by supervisors.

4. Results and Discussion

4.1. Reduction in downstream breaks

As shown by the results of the simulation, the downstream reconciliation breaks have reduced considerably by 40 to 60 percent due to the presence of centralized validation and certification logic in the proposed data architecture. Classical decentralized settings have the characteristic that reconciliation controls are replicated in many downstream systems which perform their own interpretation of data rules, data thresholds and timing. The result of this fragmentation is inconsistent results, detecting breakages too late and redundant remediation activities of reporting, risk and analytical sites. In contrast, the centralized validation strategy puts structural, semantic and quantitative verification at one validating locus (Authorized Data Store) and then only certified data is propagated to the consuming systems. The benefit of the early detection and resolution of

the data issues that occur before they flow to the downstream consumers is what causes the reduction of the reconciliation breaks. Record counts conciliation and balance and cross domain integrity checks help to remove key areas of failure that are traditionally revealed at the point of report generation or regulatory report filing. Through this, data at lower levels is loaded with information that has already met enterprise-wide quality standards, greatly minimizing the requirement of local changes or local overrides. This change of reactive to preventive control cuts on overheads and report shortening. Simulation situations also suggest that a uniform certification level and rigid snapshots add to the increased stability between reporting periods. Due to the fact that every consumer uses the same certified snapshot, any differences created by timing differences or different data extractions are reduced to none. The reduction in the counts of breaks is particularly noticeable in regulatory and risk reporting applications, where the intricate aggregation logic and high accuracy demands had made reconciliation failures common in the past. The downstream breaks that are reduced also have measures of quality beyond financial gains. Data teams note that they feel better about upstream data, reduction in the number of emergency remediation cycles and accountability is clearer when problems arise. Taken together they imply that centralized validation logic can not only enhance quality metrics of technical data, but operational efficiency and maturity of governance increasing the reliableness and scalability of this data's consumption across the enterprise.

4.2. Improved financial close timeliness

The display of the Authorized Data Store and certified snapshot framework caused a quantifiable change in the timeliness of financial close and close times have been reduced by about two to three business days. Traditional banking setup makes the financial close process too cumbersome with independent triplicates of finance, risk and regulatory report teamwork reconciles. Individual functions tend to ensure data integrity and validation with their respective extracts and control totals, commonly resulting in duplicated work, sequential dependencies and issue discovery at the end of the development. Such inefficiencies prolong tight deadlines and strain operations when there is a need to make essential reports. With all the reconciliation and certification operation moving upstream, the proposed architecture avoids the necessity of having several downstream teams that require the re Valley of the same data. A snapshot that is certified in the Authorized Data Store turns out to be the only accepted source of all of the close-related processes. Finance teams will be able to carry on with business with confidence that balances, postings and aggregations have already passed enterprise-wide quality standards and reconciliation standards. This eliminates waiting times that were as a result of inter team dependency chain and data validation cycles. Execution of closes also comes with time aligned and immutable snapshots which are used to execute in a faster way. All the stakeholders will be operating on the same point in time data set and therefore, there will be minimal discrepancies in terms of time lapse or when the data is received. Also, embedded reconciliation metrics give real-time access to data readiness, which allows solving the problem during the closing of the window, instead of its end-stage stages. Improved close timeliness brings about better control and auditability other than saving of time. Reduced close cycles minimise intervention of the manual procedures, minimise the chances of making errors under pressure and enhance meeting of the regulatory and in-house reporting deadlines. The two to three days of business savings seen do not just imply an efficiency saving, but also structural enhancement in the predictability and reliability of the financial close process and help financial operations become more resilient.

4.3. Operational efficiency

Pre-certified datasets and well-set ownership of data led to a significant increase in operational efficiency since a lot of redundancy in terms of data, finance and risk would go away because a lot of manual work went to waste. Conventional data operating models often have downstream consumers discovering problems in the data early in the reporting or analytical lifecycle, requiring manual investigation, one-off fixes and revisitation of synchronization between various teams. Such reactive remediation efforts take a lot of effort, operational risk, not to mention re-directing resources towards analysis work that add value to the business. The pre-certified datasets opportunity assists in this issue by ensuring that the data discharged to be consumed has already passed through standardized tests, reconciliation and quality criteria. Since certification is done in the Authorized Data Store and not by downstream teams, they no longer need to make unnecessary checks or compensating controls. This change allows analysts and reporting groups to concentrate more on interpretation and decision-making than on data cleansing and reconciliation. The output of simulations shows that most of the previously manual processes, say the fixing of missing records, fixing of balance errors or fixing mismatched definitions, are either removed or prevented sooner, when it is easier to fix these errors and less disturbing. Data ownership is also an important tool that strengthens operational efficiency by creating a clear sense of accountability over data quality and availability. With official designation of data owners in particular domains and datasets, problems can be channelled straight into the hands of those to be addressed without the long triage process or functional ambiguity. Ownership is strengthened using data contracts and service level agreements, which specify the expected performance and process of escalation. This design is enough to shorten the resolution time and avoidance of the revisiting well-known problems. Combined, the pre-certified data sets and explicit ownership make the data management process more than a reactive process which has little predictability and control. The workload minimization is not only effective in reduction of operational costs but also enhances the productivity of the staff, their morale and trust in the enterprise data. These efficiency improvements will create a more dependent data operating model that could serve growing regulatory and analytical needs without writing it proportional effort.

4.4. Risk and regulatory impact

The comparison of the pre-ADS and post-ADS states in terms of the percentage changes reveals the physical risk, regulatory gains realized by the Authorized Data Store pattern **(Table 1)**. Each of the metrics implies an aspect of control performance and regulatory stability and shows that the format of centralized certification and governance has a positive impact on the data control environment of the institution **(Figure 6)**.

Table 1: Risk and Regulatory Impact.

Metric	Pre-ADS (%)	Post-ADS (%)
Reconciliation Breaks	100%	30%
Close Cycle Duration	100%	65%
Manual Adjustments	100%	20%
Audit Findings	100%	40%

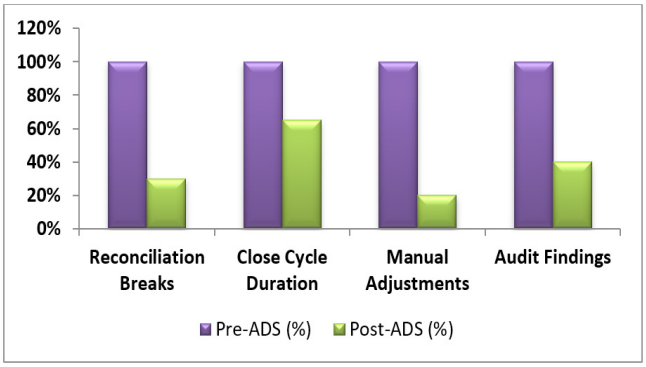

Figure 6: Risk and Regulatory Impact.

- **Reconciliation breaks:** Normalization of reconciliation breaks in the pre-ADS environment has a normalization of 100 percent, meaning the common inconsistencies among downstream systems as a result of fragmented validation logic, time warps. After the implementation of ADS, the breaks in the form of reconciliation are minimized by about 30 percent of the initial amount. This enhancement is based on the structural, semantic and quantitative consonancies that are centrally carried out in upstream such that only valid and certified information is broadcast. The reduction minimizes operational risk and reduces greatly the last-minute solutions of issues during regulatory and financial reporting the cycles.
- **Close cycle duration:** The close cycle duration indicates a decrease in normalized close cycle of 100 percent to 65 percent after ADS, which means that the timeliness significantly increased. Dependence in the finance, risk and regulatory groups have been reduced by removing unnecessary downstream reconciliations and making all consumers use the same certified snapshot. This fastens consolidations, slows delays due to data disputes and enhances the institutions response of its regulations by the submission deadline.
- **Manual adjustments:** The proportion of manual adjustments reduces significantly by dropping well down to 20 percent of 100 percent in the pre-ADS state to the post-ADS one. Pre-certified datasets minimize the requirements to make manual corrections, overrides and spreadsheet-driven fixes to address the data quality flaws in the past. The effect of this reduction is increased control performance, reduced possibility of human error and increased audit confidence in reported figures.
- **Audit findings:** Audit results are decreased by 40 percent of a normalized auditing result of 100 percent under ADS. Better data lineage, snapshot is immutable and attestation that is documented give auditors obvious repeatable evidence of data controls. This minimizes the repetition of the results on data quality, traceability and governance, proving to be more appropriate to the expectations of the regulatory environment and a more advanced data risk management process.

5. Conclusion

The paper reveals that a post-merger Canonical enterprise data model (EDM) when implemented as an Authorized Data Store (ADS) pattern can offer a structurally valid and manageable solution to most of the data integration difficulties experienced by banks after mergers and acquisitions. Traditional post-merger integration methods have been subjected to the tendency to move quickly to merge systems or even point to point data harmonization where fragmented semantics, duplicated reconciliation logic and accountability around data quality is limited. To overcome these flaws, the proposed framework proposes canonical semantic grounding and relating it to an operationally implementable data consumption layer that is the unique certified point of analytical, financial and regulatory use cases. The framework eliminates the need to handle certification and attestation conceptually, by ensuring that data certification and attestation can be directly applied to data lifecycle data, allowing it to be controlled and audited to measure and observe. Cryptographically verified snapshots at designated intervals are guaranteed to be reproducible and objective and quantitative certification scores and reconciliation measures give objective indications of data fitness. Improved accountability is also supported through introduction of explicit data contracts and service level agreements by explicitly laying out expectations between data producers and consumers. Collectively, these mechanisms allow changing enterprise data not just into an operational risk latent but into an operational asset with explicit ownership, quality targets and regulatory justification. The practical and simulated outcomes in this research indicate actual operation advantages. The reduction in the number of downstream reconciliation breaks, faster overall financial close periods and greatly reducing the number of manual reworks are all indications that centralized validation and certification substantially enhances the efficiency and stability. Regulatory wise, better audit readiness is achieved through better lineage, traceability and control evidence to mitigate occurrence of data related findings. These consequences are especially important in the situation of a post-merger, when the regulatory process becomes especially alert and the threat of data inconsistency increases due to the complexity of old systems. Although this study is based on the principles of architecture and operating models, which were mostly common before 2021, the conclusions will be very relevant. Paradigms like real-time data processing, cloud native application and distributed types of data ownership can be added as extensions to the framework, though do not supersede the fundamental requirement of canonical semantics, certified data store and formal accountability. Future studies can discuss the means in which real-time certification, policy enforcement automations and scalable cloud applications can be used to improve upon the ADS pattern. However, the major tenets described in this paper can give a long-lasting base to construct robust, reliable and regulator-compliant banking data ecosystems within post-merger settings.

6. References

1. Inmon WH, Imhoff C, Sousa R. Corporate information factory. John Wiley & Sons, 2002.
2. Kimball R, Ross M. The data warehouse toolkit: The definitive guide to dimensional modelling. John Wiley & Sons, 2013.

3. Otto B. Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. Communications of the Association for Information Systems, 2011;29: 3.

4. Hinduja S, Kunieda M. Modelling of ECM and EDM processes. CIRP Annals, 2013;62: 775-797.

5. Khatri V, Brown CV. Designing data governance. Communications of the ACM, 2010;53: 148-152.

6. Madnick SE, Wang RY, Lee YW, et al. Overview and framework for data and information quality research. Journal of data and information quality (JDIQ), 2009;1: 1-22.

7. Loshin D. Master data management. Morgan Kaufmann, 2010.

8. Panian Ž. Some practical experiences in data governance. World Academy of Science, Engineering and Technology (WASET) Academic Science Research Papers, 2010;62: 468-475.

9. Haneem F, Kama N, Taskin N, et al. Determinants of master data management adoption by local government organizations: An empirical study. International Journal of Information Management, 2019;45: 25-43.

10. Eckerson WW. Performance dashboards: measuring, monitoring and managing your business. John Wiley & Sons, 2010.

11. Power DJ. A brief history of decision support systems, 2007.

12. Davenport TH. Working knowledge: How organizations manage what they know. NewYork Harvard Business School, 1998.

13. Vom Brocke J, Rosemann M. Handbook on business process management. Heidelberg: Springer, 2010;2.

14. McKnight DH, Choudhury V, Kacmar C. Developing and validating trust measures for e-commerce: An integrative typology. Information systems research, 2002;13: 334-359.

15. Lace K, Kirikova M. Post-merger integration specific requirements engineering model. In International Conference on Business Informatics Research. Cham: Springer International Publishing, 2021: 115-129.

16. Awrey D. Unbundling banking, money and payments. Geo. LJ, 2021;110: 715.

17. Al-Fedaghi S, BehBehani M. Modeling banking processes. In 2018 International Conference on Information and Computer Technologies (ICICT), 2018: 40-46.

18. Carney JG. Industry models for enterprise data management in financial markets. IBM Journal of Research and Development, 2010;54: 6.

19. Törmer RL, Henningsson S. How enterprise architecture maturity enables post-merger IT integration. In International Conference on Business Informatics Research. Cham: Springer International Publishing, 2017: 16-30.

20. Bertino E. Systematic Control and Management of Data Integrity, Quality and Provenance for Command-and-Control Applications, 2010.