

**THE CLINE GROUP**  
Capital Markets Training & Consulting



**A New Vision  
for Enterprise Data Management**

**William F. Cline  
Founder & Managing Partner  
The Cline Group**

## Executive Summary

The current financial crisis may have been prompted by poor judgment related to sub-prime mortgages, but it also exposed a number of serious operational and information technology shortcomings across the financial services industry. On the surface, firms failed at risk management and reporting. Dig a bit deeper, and it emerges that many problems originated with firms' failure to create Enterprise Data Management (EDM) repositories.

Despite substantive investments in EDM over the past five years, few firms have created high quality, enterprise-wide reference data repositories with consistent, standardized data inputs. High-quality reference data is crucial in securities valuation, counterparty risk assessment and the maintenance of market liquidity. Yet, firms across the industry are still operating with legacy systems, complex and cumbersome centralized data models, and silos of information segregated by asset class, geography, IT and business lines.

This report examines the reasons why so many attempts by market participants to create high quality EDM repositories have fallen well short of business objectives, contributing to the current financial crisis by fostering inaccurate securities valuations and poorly grounded measurements of risk and liquidity. The few commercial EDM solutions used by the industry are mostly legacies of the past, monolithic in their design and insufficient in their use of XML and high-performance messaging to facilitate cost-effective, real-time downstream application integration. Even if reference data is aggregated, normalized, validated, and enriched in a centralized data model, it has no business value until the data is integrated into the valuation, risk management, and securities processing applications that need it.

The financial industry urgently needs to apply more modern IT and business process solutions to the EDM challenge. These centralized, legacy approaches to data management must give way to more virtual data management, distributing and caching the data required by applications closer to the point of use in the same way that distributed computing transcended mainframe computing.

Investors and regulators want, and expect, transparent markets wherein they can accurately consider the risk/reward premium associated with their investments, and this cannot be done without high quality reference data. Industry participants need to look at how analogous industry challenges such as client and risk reporting have been solved by firms like CorrectNet using more modern IT architectures, virtual versus centralized data models, high-performance messaging, and a managed services approach to unleash inherent economies of scale.

To regain investor confidence, market participants cannot continue to depend on legacy approaches to EDM, for to do so will continue the past pattern of costly and time consuming projects which nevertheless do not attain the desired business objective of higher quality reference data.



**Bill Cline**  
founder and managing  
partner  
The Cline Group

## Table of Contents

Introduction	4
The Impact of Legacy EDM Solutions	5
Centralized Versus Distributed Data Management	6
Lessons Learned From Client & Risk Reporting	7
Virtual Data Management	8
The Future of Enterprise Data Management	11

## Introduction

The global financial markets are unique in the degree of correlation between timely and high quality information and outsized profit potential. In the capital markets, milliseconds can be a competitive differentiator, while the cost of poor data quality can be virtually unlimited and put a firm out of business. These correlations are seen in a variety of different data types and use cases – market data, reference data, price and liquidity discovery, risk management, valuations, reconciliations and collateral management, regulatory compliance and securities processing.

The IT arms race is well advanced in the battle for the lowest latency, direct-from-source market data and includes in-house ticker plants, private networks, co-location of servers at execution venues, hardware and software accelerators, and high-performance messaging infrastructures. While market data is fundamental to sophisticated trading strategies, reference data comprises the securities pricing data used in position valuation, risk management and securities processing plus detailed information on a market participant's customers and counterparties. Unfortunately, current approaches to reference data aggregation, validation and cleansing, and the macro discipline of Enterprise Data Management (EDM) – where data quality and standards are more important than speed of data delivery – are mired in legacy architectural approaches and solutions not conducive to innovation.

**While “centralized” sounds like a logical approach for an “enterprise” solution, it is, in reality, inherently flawed.**

The financial markets industry has a legacy of asset class, geographic, IT and organizational silos. This legacy of silos is understandable given the fast-changing nature of the capital markets and the iterative evolution of front, middle, and back office systems and processes over the years within those silos. If it were easy to break down such silos, it would have been done a long time ago. While vendor marketing materials now promote “multi-asset class” and “multi-strategy” solutions, more detailed analysis typically reveals substantive gaps in workflow support, data models and loaders, and business rules in classes such as OTC derivatives. As such, practical data management solutions need to be architected to derive cost-effective, near-term business value by using methods and solutions that can transcend these multiple silos.

Most current data management IT solutions are grounded in the concept of a centralized data model. While “centralized” sounds like a logical approach for an “enterprise” solution, it is, in reality, inherently flawed. While a frequently used term in the industry, the very concept of a reference data “golden copy” is debatable at best. There can be legitimate business reasons why different business units may choose, for example, to use different “indicative” prices for non-exchange traded instruments or different ratings for a given security.

Redundancy in industry data identifiers is also widespread, with no solution in sight. Centralized approaches make it far more difficult to support downstream applications in different business units which use different identifiers for the same instruments. Cross referencing is inconsistent at best, and there can also be too few identifiers for classes such as loans where there is typically no standardized identifier. Conceptually, a best-practice data management solution would minimize the downstream data dependency impact on consuming applications, saving cost and streamlining time to value. It is also a politically expedient solution to support a degree of choice, as resistance to change can be well

entrenched in capital markets as in other industries. Sadly, the converse is true today in most attempts to create enterprise-wide data management solutions and programs.

## The Impact of Legacy EDM Solutions

Perhaps you've heard the old adage that "if you're a hammer, everything looks like a nail." Nowhere is this more evident than in an examination of legacy vendors of Enterprise Data Management (EDM) solutions. Having evolved over the years as their data models were extended, enhanced and customized, these centralized models became complex and cumbersome to work with, exacerbating both up-front challenges and costs of data mapping and normalization across different sources and requiring ongoing maintenance of the data models themselves to support fast-changing business needs. Vendors market the approaches and solutions they have to sell, so if you're a hammer you like to find nails.



BATS was able to design and build one of the fastest execution venues in the world by having the luxury of starting from scratch versus the more evolutionary approach at traditional exchanges made necessary by the need to support a legacy infrastructure. Few IT professionals would design EDM solutions today in the manner promoted by the legacy vendors of these solutions. The project phasing and frameworks necessary to fit with legacy vendor capabilities are one of the major reasons why so many EDM projects in the financial markets have found mixed success at best. The extensive up-front data mapping, normalization and the challenges of downstream application integration combined to ensure that many EDM projects lost executive support and budgets long before anticipated business benefits were attained. In fact, some EDM programs have been abandoned even before reaching consensus on common identifiers, business rules, or values. Other reasons contributing to project failure include not framing EDM program objectives in business contextual terms, big bang approaches versus iterative improvements, and not securing the understanding and support of business users early in the cycle.

While serving as the global managing partner of the financial markets industry practice at Accenture, I recognized the inherent industry-wide duplication of effort in reference data management. There is a finite universe of perhaps 3,000,000 securities reported on by a relatively small number of data providers and aggregators, and yet every market participant spends significant amounts of time and money to cleanse, validate, enrich, and disseminate subsets of exactly the same universe of data. IT solutions were expensive to buy and even more expensive to implement, and few market participants cleanse more than 50,000 or so securities with anything but simple, preference-based business rules. A high quality enterprise data repository was a core need for market participants but not a core competency. I saw the need for a managed service that could unleash the inherent economies of scale, so market participants could attain a far higher degree of data quality for far larger numbers of securities. This would have allowed market participants to focus on downstream application integration, the only point at which real business value is derived.

While the vision was sound and we did many things right, we also made mistakes and were perhaps too early to the market. While every other data vendor was willing to be part of a managed service solution,

one vendor – deemed a must-have by some market participants – was ultimately not willing to participate. More importantly, we also encountered first hand as a service provider the complex, expensive and time consuming task of implementing legacy EDM IT solutions. It was clear that if Accenture was having such challenges with data normalization, data mapping, and the like given a business plan grounded in one-to-many economies of scale, individual market participants were likely drowning in costly EDM programs that were slow to derive tangible business benefits.

I thought there had to be a better way, so after resigning from Accenture to start The Cline Group I began rethinking the entire approach to Enterprise Data Management. I looked for other IT and managed services providers in other industry segments that used more modern techniques to aggregate, enrich, and deliver higher quality data for more streamlined downstream access – thus speeding time to value while reducing cost. I found such solution providers in managed services for client and risk reporting – but more on that later.

## Centralized Versus Distributed Data Management

This one-size-fits-all approach of centralized data models limits flexibility to respond with solutions to legitimate business needs. The monolithic data model becomes ever more unwieldy, and data dependencies and relationships create unanticipated downstream problems. When downstream systems must adapt to a compromise central solution, changes are made downstream within their own silos so conceptual benefits of centralization are lost nonetheless and more incremental effort is expended. The number of central database queries also grows exponentially over time, which strains both capacity and performance. This contributes to latency, not traditionally seen as a big problem for referential data but becoming more so as business objectives such as real-time risk management and intra-day exception management become more urgent business needs.

Centralized approaches to data management can also create a single point of failure for the enterprise, with multiple downstream systems now dependent on the central repository for key data inputs in mission-critical applications. While some market participants have implemented hot failover EDM solutions, this can be a costly approach relative to more modern and distributed architectural designs. Furthermore, centralized data repositories can become enormous as more types of pricing, trade, quote, rating, and client/entity data are retained for longer periods of time, catalyzed by new regulatory mandates such as Reg NMS and MiFID which require broader data storage for longer durations.

**A key business driver for more distributed data management solutions is the increasingly apparent need for more real-time risk management.**

If centralized approaches are flawed, a more distributed data management approach seems like a concept worth exploring, analogous to mainframe versus distributed computing. Ironically, the IT investments made by market participants to support high performance electronic and algorithmic trading are quite similar to infrastructure prerequisites for more distributed data management: a high speed messaging bus, distributed data caches, Complex Event Processing (CEP) engines, and historical databases. Even for those still in the early stages of adoption, the imperative is clear as electronic and

algorithmic trading as a percentage of trade volume has been consistently rising and is projected to continue rising across all asset classes.

Distributed caching is perhaps the most fundamental component of distributed data management, storing the referential data in memory while distributing and duplicating it throughout the enterprise. Metadata is resident in the data cache and is accessible via an Application Programming Interface (API) or other means. In this scenario, one distributed cache can send its updates to other caches aligned with other business units or geographies. This proliferation of more localized copies puts the source of data closer to the business need in order to reduce latency.

High speed messaging infrastructure works in concert with the distributed caches to parse data at the source closest to the data, with subsets distributed to other subscribing applications from that point. This more distributed data model can coexist with legacy security master vendor solutions, and it makes the location of the data moot for subscribing applications.

It should be noted that distributed data architectures also push data governance down to the application level to simplify access to disparate data sources, away from the database itself. The instantiation of objects occurs in the cache so it understands how to deliver the data. This creates flexibility by eliminating direct dependencies between consuming applications and a “golden” security master, and it sets the stage for distributed reference data caches tailored to meet specific business objectives. Best of all, this all happens dynamically and distributes the processing load much more broadly than any centralized approach can provide – thus fostering a lower latency solution.

A key business driver for more distributed data management solutions is the increasingly apparent need for more real-time risk management. For static buy and hold portfolios, end-of-day batch processing and file-based distribution of securities pricing is adequate. However, these end-of-day cycles are inadequate to measure fast-changing positions and risk profiles associated with high frequency trading. Since the pace of trading is accelerating with no apparent end in sight, it is logical to assume that support for real-time applications will factor more substantively in charting the optimal approach for EDM programs. Distributed and more virtual data management architectures put data repositories in much closer proximity to the points of usage, and complete data sets do not have to be distributed to every node on a network in favor of dynamic caches, thus eliminating this expensive and unnecessary network overhead.



## Lessons Learned From Client & Risk Reporting

In my search, I found that the company that best embodies principles of a more modern approach to data aggregation, normalization, validation, enrichment, and distribution is Long Island, N.Y.-based CorrectNet. CorrectNet is a managed services provider of client and risk reporting for global market participants. Their clients include many of the top global banks, two of the top three prime brokers, eight of the top 12 global fund administrators, 42 of the top 200 hedge funds, and over 400 independent Registered Investment Advisors (RIAs). One of the very few successful onshore domestic U.S. BPO providers, CorrectNet

operates over 600 private-label portals on behalf of their clients, with over 500,000 end users accessing CorrectNet managed data and reports daily in over 2,000,000 files.

When one understands the challenges inherent in legacy approaches and solutions to the EDM challenge, it is easy to see how more modern approaches used in client and risk reporting could transform the way the financial services industry plans and executes EDM programs. First, let's contrast what CorrectNet does with the myriad data feed, file inputs or data extracts they aggregate and validate for reporting purposes. Rather than using a rigid, centralized data model to which everything must be mapped, CorrectNet employs a virtual data services architecture with little dependence on where the requisite data resides. It's all "in the cloud" so to speak, with some data being hierarchal, some regional, and some aligned specifically with a line of business. Data as a service – sound familiar?

CorrectNet applied modern approaches for a data integration solution and then combined it with managed services and scale that made their solution more cost effective and flexible than any in-house reporting alternative. The smart use of IT enables CorrectNet to compete as an outsourced client and risk reporting managed service provider without offshore labor arbitrage as any component of their value proposition. Their solution is architected to allow for easy enhancements to functional "tiers," so changes and improvements can be made without impacting the broader infrastructure – in contrast to more monolithic legacy designs.

In contrast to expensive and time consuming up-front data mapping, the CorrectNet managed service infrastructure provides for automated data transformation, discovery, and schema inference. In other words, once the file is loaded you can visualize it immediately. From there, they tune the schema for various data types, domains, names, and links as part of the overall data normalization process. Clients can select from predefined and configurable enrichments for aggregation, delta, heat mapping, and more, all of which are designed to facilitate alternative views, formats, and distribution methods – thus minimizing impacts such as programming effort on downstream consuming systems.



## Virtual Data Management

Client reporting typically has to aggregate an equally if not more expansive universe of data types and sources than an EDM program, so one might look at this discipline to illustrate more virtual and distributed methods that could be used to address the industry's data management and reference data quality challenges. In contrast to the monolithic, centralized approach and the hammer and nail analogy, let's look at how data normalization could be done with more distributed and web services grounded techniques.

To support disparate content formats and frequently overlapping data identifiers, CorrectNet employs a "virtual data model" that dynamically structures data. It is grounded in the following high-level process flow:

- **Standardize all data to XML**
- **Normalize to required data structure and schema**
- **Link to master files (e.g. data identifiers) via cross reference services**
- **Apply rules-based exception processing**
- **Enrich data according to service need (aggregate, roll up, etc.)**

The virtualization of data content means that there is no centralized data model or central data repository. Instead, unique “views” are created that support business specific requirements as to format and means of access. In this manner, all data becomes “self describing” in nature and fosters more distributed data management, distributed data caching, and enables the data to conform to the application’s format expectations as opposed to forcing myriad applications to conform to a centralized data model.

Consider the scenario of multiple bond ratings and a business unit’s desire to have a waterfall decision tree to decide which ratings to use in different regions and for different asset types. Downstream systems may also want different outputs from a waterfall tree; one group may want a rating hierarchy that included internally derived ratings, while another group may only want to consider agency ratings. These use cases put further pressure on the very concept of a “golden copy” being the ultimate EDM program objective. For client reporting, CorrectNet uses XSLT business rules which allow these examples to be easily accommodated. The role of XSLT is to transform XML data from one representation to another, and XSLT can separate the transformation into two streams – one for applying the business rules and another for formatting.

CorrectNet’s virtual data model and aggregation techniques allow them to easily process any type of file or feed – even unstructured data – and supports web service messaging both pushed to them, and called by them. Data can also be extracted if it is a web-enabled service, and for application integration they have a sophisticated API and publish every view and data set as a Web Services Description Language (WSDL) callable from any web services infrastructure. WSDL is an XML format that describes network services as a set of end points, with messages and operations described abstractly and then appended to a network protocol and message format. WSDL is extensible in that it allows descriptions of messages and end points agnostic as to the message format or network protocol.

**Most files are still sent end of day, one reason why the elusive goal of real-time enterprise risk management remains far from the norm in an industry where low latency solutions and front office technologies are so advanced.**

Boston research firm Aite Group expects \$2.5 billion was spent in 2008 addressing the EDM downstream data integration challenge, a 20% CAGR to serve the many business requirements. Research also indicates that almost 40% of market participants have between 21 to well over 100 downstream application connections per firm. Sadly, specialized EDM technology vendors have only a few off-the-shelf connectors to support this diverse multitude of downstream consuming applications. The result is

expensive and time consuming custom development for downstream connectivity. This further extends time to value for the typical EDM project and is a major reason why so many EDM programs over the past few years lost political and budgetary support long before the promised business benefits were attained. The number of custom connectors also contributes to the ongoing maintenance challenge, limiting flexibility to adapt to changing business needs and threatening compatibility with future software releases.

Here too, a legacy way of thinking and architecting solutions for downstream connectivity exacerbates an already challenging situation. The Aite Group estimates that over 60% of downstream EDM connectivity is comprised of file-based transfers. It is rather curious that in 2008, in the most real-time industry in the world, batch programs generating file outputs were still the most common means of feeding legacy systems and sub-security masters. Most files are still sent end of day, one reason why the elusive goal of real-time enterprise risk management remains far from the norm in an industry where low latency solutions and front office technologies are so advanced.



If solutions are to evolve to foster more streamlined and real-time downstream connectivity, it is clear that more modern approaches are needed here as well. Once again, one might look to client reporting and a managed services provider like CorrectNet for lessons that could be applied to EDM. CorrectNet's subscription services approach configures a "view" of data which is actually a content structure, allowing for the use of "publish" and "subscribe" in the same manner used for real-time market data distribution to the trading floors of the world. If EDM infrastructures are ever to support real-time or even intra-day business requirements efficiently, there are few alternatives to such a "push" approach coupled with a robust entitlements schema.

It seems certain that if specialized EDM vendors had the luxury of designing a solution from scratch today, this more distributed and virtualized web service enabled approach would be the path of choice. I am not stating that client reporting managed service providers like CorrectNet will definitively expand their strategy and services for EDM, but months of study have convinced me that their innovative techniques could well be applied to the industry's EDM challenge. I view CorrectNet's use of xml, distributed caching, virtualized data models, and managed services as complementary to legacy EDM approaches and envision a way to "wrap" legacy EDM-centralized solutions – most of which are only partially implemented – to create a more flexible and distributed EDM infrastructure that speeds time to value.

In this paper, I have tried to explain the limitations of legacy EDM solutions and traditional program approaches, and to draw a correlation to another discipline within capital markets that faces similar challenges, yet has transcended them with more modern methods and processes. The challenge the industry faces for improved data quality in a timely and cost effective manner is sufficiently urgent that it seems perfectly reasonable that two vendors coming from different origins may combine forces to deliver real business value to an industry under pressure.

## The Future of Enterprise Data Management

Most market participants rightly see EDM as a core need but not necessarily a core competency. Perhaps we were early in our attempts to unleash these inherent cross-industry EDM economies of scale at Accenture via a managed service, but the duplication of effort in the data management challenge still persists across all market participants and geographies. At a time when market pressures are forcing everyone to look for ways to reduce costs while improving transparency, the data quality imperative is even more urgent and evident today than ever before. I feel sure that lessons learned from client and risk reporting will be applied to a new generation of EDM solutions and managed services to finally solve, in a reliable and cost-effective manner, the ongoing reference data challenge.

**It is clear that static and ineffective market and credit risk management, poorly grounded valuations, IT and organizational silos, and inconsistent reference data quality were all contributors to the current financial crisis we now face.**

Despite many years and billions of dollars of investment, it is clear that static and ineffective market and credit risk management, poorly grounded valuations, IT and organizational silos, and inconsistent reference data quality were all contributors to the current financial crisis we now face. If capital market participants and solution providers do not coalesce to apply more modern approaches to EDM on which improvement in each of those contributory factors depends, it is likely that regulatory authorities will intervene in ways that few market participants would welcome. It's time for a new way of thinking about Enterprise Data Management and referential data quality to meet the current and rapidly evolving needs of the capital markets industry.

---

*Bill Cline is the founder and managing partner of financial markets training, recruiting, and consulting firm The Cline Group ([www.billcline.com](http://www.billcline.com)) and the former global managing partner of the capital markets industry practice at Accenture. Previously, as a principal at product firms Rich Inc. and FD Consulting, he introduced a number of significant innovations to the industry including distributed messaging platform Triarch and the first commercially available ticker plant TIPS.*