Data Issues for Analytics

CDO 2.0: Managing the New "Data-to-Value" Equation

July 2018

Michael Atkin, EDM Council, Strategic Advisor

Data Issues for Analytics
CDO 2.0: Managing the New "Data-to-Value" Equation

The CDO is being challenged to step up to a new mandate for data. In addition to supporting the "control environment" needed for operational data management, today's CDO must also be thinking about managing data to support innovation. And managing data for "insight" is fundamentally different than managing data for "operations." Both are important to the firm. Operational data management is grounded in precision of meaning. The operative concepts are about production rigor, ownership, access control, lineage and data flow. Managing data for insight is more aligned with agility, self-service and processing capability. From this perspective, data is viewed as a social asset with emphasis on what the data represents, how it was derived and how to turn "ideas into discovery."

Data for Operations	Data for Insight
"Block and Tackle" (Risk, Regulatory Reporting, General Ledger)	**"Ideas into Discovery"** (Innovation, Speed and Self-Service)
Precision of Meaning (data quality, control mechanisms, governance, change management)	**Basis of Data** (how created, how derived, rights to use, APIs, access, IBM's 7Vs)
Lineage and Data Flow	Time to Market
Reference Data (harmonization)	**Alternative Data** (direction)
Production Rigor (accountability)	Agility (transparency)
Critical Data Elements	**Product-to-Customer Concepts**
Control Environment (ownership and stewardship)	**Social Asset** (Center of Excellence)
"Solve for Problems" (reconciliation and transformation)	**"Solve for Patterns"** (automation and processing)
Definitions and business rules to ensure harmonization	Tools and processes to model, visualize and query
Data Manufacturing Process	**Transformation and Optimization**
Business Metadata (meaning)	Technical Metadata (structure)
People Intensive (logic-based rules)	**Process Intensive** (machine learning-based rules)
Chief Risk Officer	Chief Analytics Officer

Summary of the Data Challenges Facing the CDO

Setting the Baseline

Operational data management for financial services is grounded in the ability to model legal obligations where precision and harmonization are essential. We rely on precise data to ensure

accurate projection of cash flow, unwind business relationships and manage nuance across complex financial processes. We need to share data between lines of business taking into consideration issues such as liquidity, exposure and collateral all while faced with continually changing macro-economic, political and operational events.

To accomplish these objectives, financial institutions need to aggregate this information from decentralized data repositories and extract it from a spectrum of existing systems across their enterprise. The data needs to precisely describe the terms and conditions of the totality of the financial deal at a very granular level. Data aggregation encompasses front, middle and back office applications all with varying levels of data precision needed to address the specifics of their legal obligations. The challenge for data management is that this data is extracted from a multitude of systems leveraging existing technology and modeled using proprietary schemas.

It is from this operational and legally interconnected morass that we've adopted the concepts associated with a "data control" environment. It's a rigorous process. We spend countless cycles reverse engineering our business environments. We debate the concepts of criticality needed to manufacture and replicate business outcomes. We anchor data to precise meaning and worry about accurately modelling operational reality. We invest heavily in managing the relationships between content production and data consumption. We struggle with defining authorized data domains. We document transformation rules and trace data flow. We have access restrictions for privacy and entitlement control mechanisms to track intellectual property rights. We manage data sharing requirements and define service level agreements.   And we seek to manage it all via rigorous policies with governance controls and cross-functional accountability matrices.   All this just to manage one half of our data objective.

Precision versus Time-to-Market

At the other end of the spectrum is data used for analytical insight. For this new equation, firms are looking to make the transition from "control" to "agile." The bigger issue is time-to-market. This is the currency of the financial industry. The existing approach of applying rigor to manage data meaning and quality was born for regulatory compliance with a host of laborious and time-consuming processes that impose a significant cost and accountability burdens on stakeholders.

Data management for business insight is not the same. The new digital world must deal with a host of enhanced challenges associated with processing data at scale. This can be hard to accomplish particularly with rapidly changing data types using a variety of data formats. We need to reinvent routines and processes. Legal meaning is not the same as business insight. The goal is discovery, speed and flexibility. Data management processes can't be an obstacle to innovation. The stodgy old-world of data management is a cultural mismatch for agile operations. We don't want the lines of business to run through the gauntlet of pull requests, massaging and curating the data just to gain momentary insight.

Our new goal is to put discovery into the hands of the business analyst. That means more time spent on requirements capture and capability. More time on translation in support of discovery. More time getting our data production processes in line with how the business operates. More time acquiring the data the business wants for their models.

And this puts real pressure on the Office of Data Management. Stakeholders are not keen on having to redo processes they were forced to adopt. Data fatigue is a real problem. We promised that the regulatory control environment (which is still fundamental) would be useful for business analytics. This turns out not to be the case for many applications, so now we also have a credibility problem to address. It is also true that business users are impatient. They want to easily plug data into their operational processes to link concepts to transactions to products to customers to instruments. We must learn how to move from the old model of "read and report" to the new goal of "input and process."

In a very tangible way, those of us who survived the initial wave of data management are now dealing with the need to change our platforms, modify our routines and alter our models. The pace of IT innovation and analytical capability is growing faster than the ability of the ODM to respond. And this can be challenging because humans (and large corporations) are not always good at change management or on-the-fly adjustment of operational routines.

Different Sort of Data Wrangling

Operational data management has traditionally focused on anchoring data to precision and meaning. We accomplish this objective in three broad ways.

1) Reference data (i.e. terms, conditions, classifications, descriptors, instructions, rates, dates, features, etc.) is grounded in legal obligation. These represent the contracts and agreements that result from the issuance and actions processes that propel our industry.

2) Customer and product information are contextual. These data sets are grounded in consensus across the stakeholders associated with the individual institution.

3) Transactions, pricing and holdings are point-in-time events. Precision in this domain is captured by the metadata that describes the time, amount and quantity of the event.

Data management for analytical insight is anchored in process. To define the data agenda, we must think differently about the data objective and the requirements for "what happened" (descriptive) versus "why it happened" (inquisitive) versus "what's likely to happen" (predictive) versus "so what" (prescriptive). The data management objective in this sphere is very different from the stringent quality and precision requirements for critical measures such as risk weighted average, aggregate notional value or cumulative exposure calculations.

The bigger CDO concerns are now with concepts such as content disparity, where there is a need to prepare the data for analysis. The breadth of data also becomes critical because of the increasing number of parameters associated with data obtained from multiple sources. In circumstances where the focus is on the holistic view (i.e. for historical transactions or time series analysis against political or social trends), the importance of any single factor of input can be reduced and replaced by model parameters that can be adjusted to account for quality gaps in the data.

Stakeholders need to understand the "basis of the data" including how it was created, where it originated, how it was derived and how it is being maintained. The requirement for data accuracy can be less stringent. For example, if an analyst is looking at tens of thousands of trades, the bad data becomes irrelevant because the analytical objective is for "direction, not precision." Business analysts and data scientists can build the data quality expectation problem into their models to reflect accuracy and trustworthiness requirements. As a result, today's CDO not only needs to rethink the relationship between data producer and data consumer but must become intimately familiar with the new requirements for predictive modeling (to unravel scenarios and identify patterns), advanced query (to follow an idea into discovery) and data visualization (to understand interconnections) – the big three for data analytics.

Credibility, Trust and Transparency

The key to managing data, regardless of application, starts with the implementation of a functional data infrastructure. Successful companies agree. Get the "blocking and tackling" of data management in place before you do anything else. This means embracing the concepts of meaning management, implementing governance accountability, building inventories, establishing control processes, managing metadata and establishing trust. Good data hygiene is the baseline for everything else and essential for all analytical processes.

Data that is used for general ledger or regulatory reporting must be managed to a high degree of quality. It is the price of entry and essential for the core work of the financial institution, but not very interesting to the lines of business. The key to serving the R&D side of the house is effective communication and an ability to relate to these new analytical requirements. Here's where governance credibility can be leveraged. If analysts trust that the ODM can deliver on the compulsory data, they will trust you on the innovation task.

In this new collaboration model, the relationship between data production and data consumption is still critical. The data scientist defines the requirements for their models and queries. But in addition to the control environment for precision, the CDO must also focus on format transformations to optimize for expansive quantities of data. Technical metadata becomes as critical as data meaning. The fit-for-purpose criteria is intricately linked to how the data will be used. Analytics is managed by a different process. The new goal of the CDO is to eliminate the need for transformation – the cause of most operational data problems in our

industry. Data needs to be managed closer to the line of business application. It is supported by machine learning to identify data non-conformance. It is less about the "golden copy" and more about managing the semantic data lake that also becomes the distribution hub for organizational content. It is about leveraging the common business process for dealing with data challenges.

Managing data in this unfamiliar environment is not without its challenges. There is still the need for operational control and pristine data for general ledger, risk management, legal compliance and regulatory reporting. The goal is balance between control and flexibility. Friction comes into play with new teams of users that don't have experience working with the CDO and who don't have the basis of understanding and trust in the governance framework.

One of the key areas of concern is with "data anarchy." Most of us in the data business understand this as the original data management problem. It is caused by vertical alignment of data from local repository to business application. It is this independent management of data across fragmented IT environments that enables data to easily get out of alignment. It is in this "environment of anarchy" where data is modified, transformed and renamed to match the schemas of the proprietary software and data models that drive our applications. We do this because we are focused on making the application work in our vertical process. We don't worry about harmonizing the data to precision of meaning so that it can flow horizontally across linked processes.

This problem can be magnified in R&D environments. There is a plethora of tools that are readily available, and the pace of change can lead some to forget to follow the existing control processes that we worked so hard to establish. It comes with the territory. The process of discovery that we seek to nurture frequently results in the creation of new data sets. If the analyst is adding inputs into the new data set (and onward redistributing the resulting content), how does the CDO ensure alignment with the original sources of the data. It is the availability of tools combined with the lack of governance and ineffective control processes that exacerbates the potential for end-user anarchy.

If Data is the New Oil

The answer to data anarchy is (of course) governance. One of the goals of the CDO is to empower the user and support the production of "new data." And once created, effectively designed governance processes exist to ensure that the metadata (i.e. designed by, objectives, business rules, data used, calculation methodologies, workflow) is documented and that the new data sets are added to the organizational inventory.   Analytical data can become a true social asset of the organization and supported by newly created centers of excellence. Transparency about the production process facilitates shared learning. Analysts with linked interests can be connected. Time to market can be improved. Commentary on the value and usefulness of the new data sets can be added as part of the inventory.

Data must be understood as a factor of input into the knowledge production process. Content for operations and data for analytics should be viewed as two sides of the same coin. Some data must be right. This is data that moves through our operational infrastructure and is needed for efficient business process execution. Data for insight must be grounded in a fit-for-purpose framework designed around the needs of the business analyst. Both sides can be managed with a wisely constructed governance process. This is a cautionary tale. The CDO who is only focused on the traditional (operational) role might find oneself left behind as antiquated and irrelevant. One of the CDO's stated it precisely … "if data is the new oil, sometimes you need a different refinement process for the final product."

EDMConnect

JOIN THE CONVERSATION!!!

To be a part of the discussion, we invite you to comment and discuss the concepts and points discussed in this paper via the new EDM Council's collaboration platform, EDMConnect.
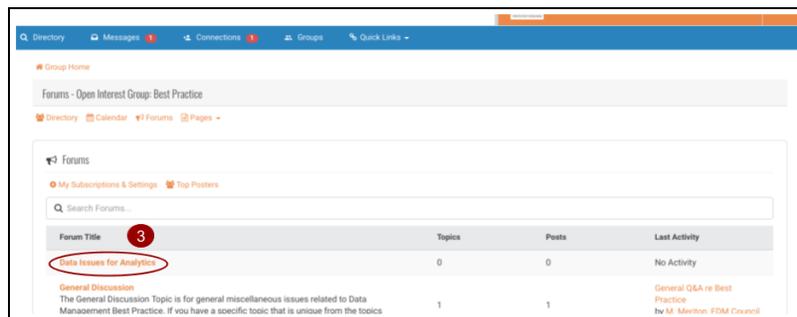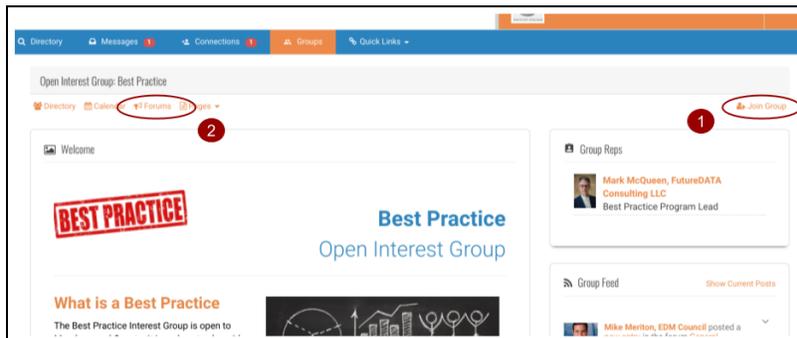
Sign-in

Logging into EDMConnect depends upon your platform status:

- If you have already activated your EDMConnect registration, select – Login
- If you have an EDM Council login, but have not yet registered, select – Reset Password
- If you are NEW to EDMConnect, please select – Register

Four Easy Steps - Go to Best Practice Interest Group

Click Here

1. Join the Group
2. Select Forums
3. Select Topic: Data Issues for Analytics
4. Post your thoughts and comments

While in EDMConnect, explore other forums and groups to find many issues of interest to Data Management professionals like yourself.

Collaborate, Contribute and stay Connected!