

## **Data Quality is the Objective**

Michael Atkin, Managing Director, EDM Council Inc

February 26, 2007

Most everyone is aware of the changes that are taking place throughout the financial industry. Business requirements have changed. Customers are more sophisticated. Regulators are paying attention to the notions of transparency and reporting accuracy. Products and markets are more complex. And margins continue to be squeezed along with cycles. Equally obvious is that in order to operate in this environment firms need accurate, consistent, transparent and precise data. They need it to support innovation, meet regulatory obligations, manage risks, automate processes, serve customers and operate efficiently.

But the purpose of this article is not to simply reinforce the core hypothesis (i.e. that the management of data content matters). Instead we believe the financial industry is at a rare and wonderful point in its lifecycle called opportunity ... to achieve more confidence and trust in data ... to automate business processes ... and to specify new requirements to both data and platform vendors.

The one thing I've come to appreciate as it relates to data quality is the importance of focusing on the root cause of challenges rather than on their symptoms. The root path to improved data quality begins with (1) getting the core identification of instruments, legal entities and data elements standardized and implemented, (2) tagging essential data elements (using the above standards) at the point of data origination (i.e. with issuers) rather than downstream and (3) taking advantage of the opportunity to partner with data vendors to improve the data manufacturing process rather than only focusing on fixing data errors when they show up in downstream applications.

The bottom line is that these types of opportunities are few and far between - and now that the light is shining on data quality and content management -- we don't want to waste it.

### **Opportunity One: Get the Data Infrastructure Correct**

Unique and precise identification of financial instruments, data terms and definitions (ontology) and legal entities are at the foundation of effective data management. If the foundation is weak, the quality of the entire 'data building' is at risk. Below is a quick snapshot of the activity associated with the precise identification of stuff:

#### ***Financial Instrument Identification***

Financial instrument identification provides the historical context. Least we forget, in 1968 the financial industry was drowning in paper and reeling from manual reconciliation. It was on the back of that 'paper crisis' that a working group of industry executives championed the use of standards for identifying securities (known as the Committee on Uniform Securities Identification Protocols - CUSIP) to promote automated processing of transactions. Notwithstanding the fact that it took over 20 years for CUSIP to be widely adopted (and ignoring for this article the challenges associated with multiple listings) - standard securities identifiers are a requisite for doing business.

**Path forward:** The industry has done/is doing the right things as it relates to instrument identification. Back when T+1 looked like a regulatory mandate there were big and significant problems associated with assignment and maintenance of security identifiers. ANNA and the ANNA Service Bureau stepped up to the plate and addressed those gaps - and should be congratulated (even though some difficult commercial challenges remain) for their efforts.

### *Legal Entity Identification*

I suggest we are facing a parallel circumstance with legal entity identification. ISO standards exist for settlement and clearing agencies, custodians, exchanges/trading systems and fund managers but not for issuers, corporate clients, suppliers and funds (together referred to as legal entities). And they are clearly needed ... for counterparty and issuer risk management ... for regulatory reporting ... for corporate research ... for legal agreements and documentation ... and for counterparty identification on transactions. Not only has the requirement been validated, the basis for the legal entity standard (i.e. format, structure, assignment process, maintenance, QA assurance) has been articulated - since 2004!

So what's the problem? In short - it is both the commercial model and the bureaucracy governing its implementation. This issue is at a stalemate. We want ISO, rather than a proprietary identifier, to be the standard and yet ISO appears unwilling to advance the discussion because to do so requires ISO to think outside the limitations of its existing commercial model. Assigning identifiers to supply chain participants and managing implementation across hundreds of involved parties requires investment. No company that is capable of managing both assignment and implementation (and there are many that fit the criteria) will step up to this investment challenge without a clear ROI. We all know that nothing worthwhile gets done for free - just as we all know that cost-based recovery models don't work. Core data identification standards must be viewed as a business proposition if we want to attract quality companies into the fray.

**Path forward:** ISO has been remiss by not stepping up to the underlying commercial framework challenge associated with legal entity implementation. And I'm not advocating a prescriptive model for the standards commercial structure. Instead, I believe the stakeholders (i.e. the financial institutions) must be willing and encouraged to set out the requirements as well as the business ground rules. Get the requirements correct and let those inclined to participate apply their commercial creativity before the fact, not after the horse has left the barn.

### *Data Ontology*

There are hundreds of data attributes associated with financial instruments for set-up, pricing and corporate actions maintenance. And there are numerous sources for that data. Integrating multiple sources into internal and external applications is difficult in itself. The problems are compounded with the lack of standard nomenclature and imprecise data definitions. Complex analytics, new source addition, automated models and data interchange throughout the transactions chain (internal and external)

all require precision. And precision is best built on the back of standard data terms, definitions and core business relationships (TDR).

The industry understands this requirement and has been working in fits and starts to standardize TDR since the turn of the millennium. And I don't want to minimize the progress that is being achieved. In fact, I'm encouraged by it. The problem with achieving this objective is first to get clarity and broad buy-in to the objective itself - which I believe is a consistent and well maintained business ontology (not an XML schema (i.e. MDDL) or a data model (i.e. ISO 19312). And while schemas and data models are good things, they are not the same as a precise requirements statement to define the business TDR - from which all data sources/feeds can be compared against and mapped to.

**Path forward:** I propose that data ontology is the starting point. ISO - you do not yet have consensus on 19312 and you must get it before you try to implement a standard if you want to serve the requirements of the financial institutions and get the standard implemented in our lifetime. Most of the firms I work with have no real clue about either the objective or the process associated with this critical activity.

The good news is that the firms are paying real attention to the importance of data precision. The additional good news is that US firms are currently engaged with DTCC on a similar activity. These are mutually supportive activities. Now is the time to reinforce the mandate - not after the fact. I encourage you to stop trying to rush headlong into completion of the 'data model' until you get alignment (at the right levels) from the industry on what they really want from this activity including how it will be actually used and the requirements associated with its implementation and ongoing maintenance.

### **Opportunity Two: Markup at Origination**

One of the most appealing pathways toward improving data quality is the concept of standard markup at the point of data origination. The logic is straightforward. Many of the data attributes that find themselves into masterfiles are pulled directly from the legal documents associated with bringing new issues to market, regulatory filings and corporate actions announcements (i.e. official statements, prospectuses, private placement and offering memos, terms sheets, trust agreements, indentures, proxy statements, CUSIP filings, ordinance documents, series declarations, etc.)

If issuers tagged these documents with the standard TDR at origination there would be fewer problems with field translations, inconsistent symbology and mapping. The trend toward electronic regulatory reporting combined with the progress of XBRL and the potential of the ISO 19312 initiative gives me some small degree of optimism that this is more than just a pipe dream.

### **Opportunity Three: More Transparent Data Manufacturing Processes**

Financial information is a manufactured product with vendors (i.e. Reuters, IDC, Telekurs, Bloomberg, S&P, etc.) still acting as the primary manufacturing agents for financial institutions. The current focus on data precision is propelling many firms to work more closely with these vendors to better understand (and make more

transparent) the processes they employ as part of that manufacturing process. This new focus on the data creation process is helping to raise the bar on data quality and is prompting vendors to pay attention to data utilization rather than just bulk delivery.

We think this is good news for the industry at large. More transparency on data manufacturing shifts the discussion away from traditional definitions of added-value (i.e. acquisition, symbology, normalization and distribution) which are becoming more commoditized and pushes vendors in the direction of hard to get data, analytics, integration into applications and contextual content (i.e. the next level of added value). And after all, overcoming data scarcity, paying more attention to how data is being used and adding secret sauce to make data more functional is what financial institutions want from their data vendors.

## Summary

From our perspective, it is time to put all of the attention and activity on data quality into proper context. The goal is to improve the efficiency of the financial information supply chain. Standards for unique and precise identification are an essential prerequisite. Adoption of those standards by vendors and data originators are where the payoff gets articulated. The rising tide should ideally lift the entire data quality ship and help propel the industry toward the next levels of added value. Time to get on with it.

*Michael Atkin is Managing Director of the EDM Council a non-profit trade association addressing the business strategies and practical implementation realities of enterprise data management. Comments, suggestions, rants and raves can be directed to [atkin@edmcouncil.org](mailto:atkin@edmcouncil.org)*