DM Review

Turning Data Into Intelligence

February 2005 www.dmreview.com

data stewardship

The Case for Data Stewardship

By William Laurent

The Case for Data Stewardship

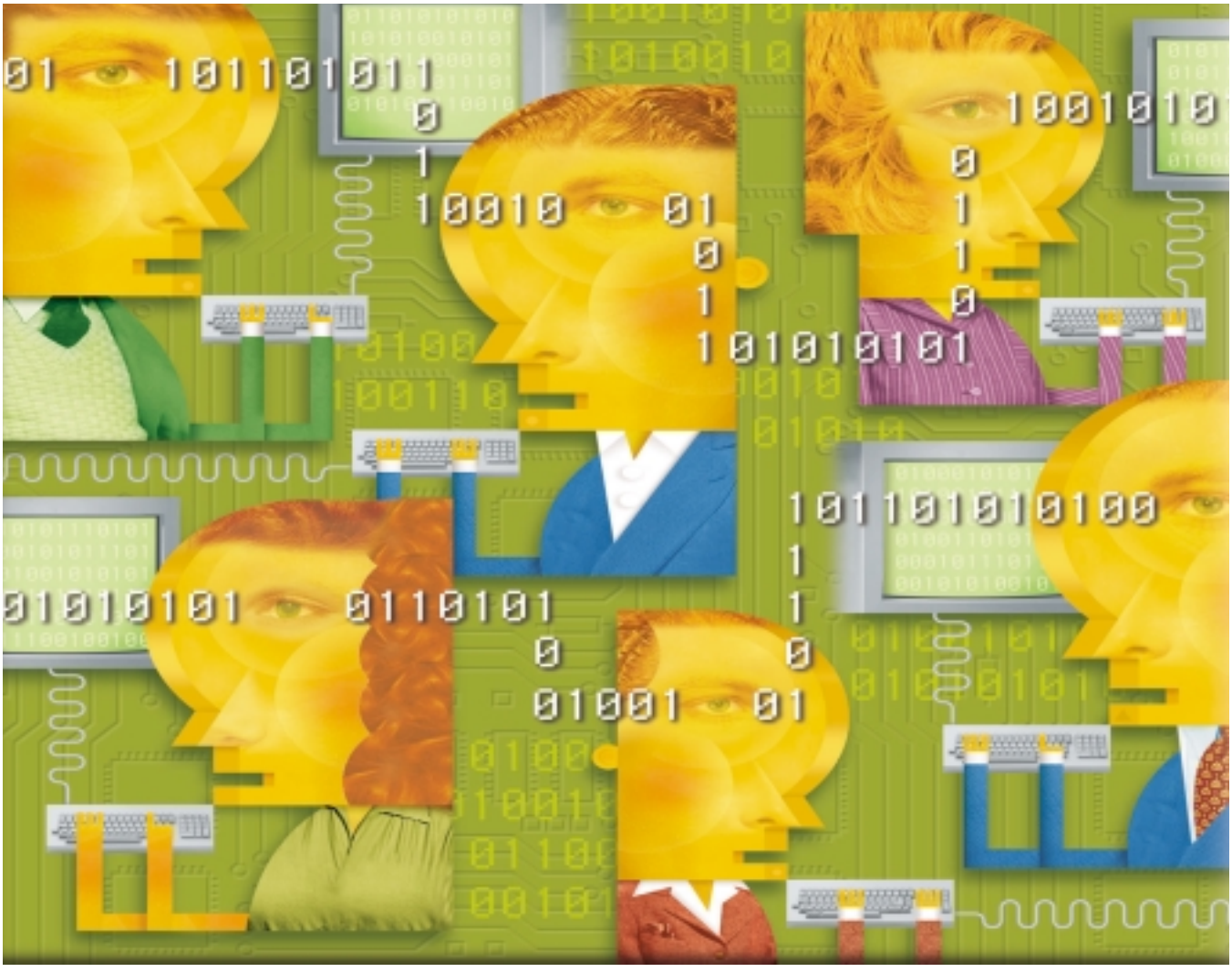
By William Laurent

The Sarbanes-Oxley Act of 2002 (particularly Section 404), has created a sense of urgency for senior executives to take an active interest in the accuracy, consistency and timeliness of their data. Now more than ever, corporate compliance, audit and governance issues must be efficiently resolved. Yet it is widely believed that less than half of large companies have a formal data stewardship or data quality program that protects and leverages their unique strategic asset of data. Smart businesses are proactive, not reactive; they comprehend the business need for quality data. Aside from compliance and governance agendas, having first-class data (and the ability to unlock it) significantly helps all aspects of your organization's business – helping to effectively gauge and manage risk, avoid redundant work loads, increase customer satisfaction (on the front and back ends) and provide for better business intelligence and decision making support. Conversely, poor data quality will have injurious consequences across all enterprise operations – negatively impacting the most critical customer service obligations (such as billing and remittance processing), skewing decisions and fulfillment obligations, and creating a host of other tactical problems. Numerous hidden business costs such as revenue leakage and higher capital expenditures may all be maladies that stem from lack of data integrity. While there are tools that can help with and augment your formal methodology, capable data stewardship requires the symbiotic merging and integration of the automated (technology) with the manual (people). For many organizations, data problems remain secreted or hidden until issues are uncovered that result from a formidable event of costly business impact, such as an external audit, fraudulent activity, subpoena or valued customer satisfaction issue. A prolific variety of complex legacy data (which may go through many iterations of manual reconciliation on its way up to senior management's eyes) lingering in cloaked and "siloed" structures will always be poised to wreak havoc when least expected. Today, with the advent of widespread Internet and e-commerce applications, the data problems of an enterprise can be exposed to the entire consumer world. If you have bad data on your customers (e.g., not recognizing them across accounts), they will know – often long before you catch on. The time to establish a formal data stewardship program is now!

When implementing a formal stewardship program or policy for your organization, it will be important to appoint the right person as the lead steward and assign the steward an appropriate title. This person will understand and approach duties from the perspective that data quality is a collaborative business and IT matter as he/she champions high-quality data across all multinational systems – from repositories to reports. The most effective stewards will be familiar with core business values and practices but should also be able to understand data models, tech-speak and data storage topics from a high and low level. (A systems analyst background with relevant industry understanding and outstanding communication skills will be ideally suited to bridge this sometimes giant gap between technology and the business.) A strong leader and a people person, the lead steward will educate and expand many people's horizons about proper data governance and the consequences of unreliable data on business objectives; the steward will make others accountable for continuous improvements in the caliber of data. Lead stewards should be visible senior-level people who are respected and well liked in the organization, with the ability to motivate and envision change from a high level. They should be empowered by senior management and steering committees to directly

address issues and manage standards-based implementations from both a business and technology-centric view, brandishing their "data police badge" when staff members resist data standards and the added responsibility or loss of control that come with such regulations. A VP-level title of data steward is not out of the question.

Stewards must have specific and measurable goals for data quality, making sure that public data helps enforce and promulgate vital business rules and processes. A viable formal stewardship policy will be rooted in ongoing standards that identify goals, priorities and quality metrics within all systems infrastructure elements (from data warehouses to OLTP applications) and business functions that touch or affect data. The steward will tie business strategy to data strategy, applying generally accepted qualitative metrics and heuristics (risk management, cost benefit analysis, change management, etc.) to the measuring and enforcement of data quality. With increasingly aggressive life cycle timelines and rapid application development (RAD) paradigms, organizations should not have to keep worrying about the veracity of data from conversion to conversion, re-inventorying and re-reconciling data elements piecemeal every time a new systems integration project is tackled. Without ongoing business driven tools and touchstones for data assessment and improvement, your data stewardship program is doomed to fail. Only a continuous and structured methodology will bring long-term benefits, not a sustained culture of quick data fixes. For larger enterprises, this methodology will have functions that leverage time-honored calculations, derivations curves and benchmarks for measuring quality. The business rules of the corporation should be documented and maintained alongside data models, data dictionaries, meta data repositories, etc., so that a spirit of common responsibility, knowledge and ownership is cultivated. The formal stewardship policy will reside in a public place (corporate intranet) along with other operating policies so that employees who create, use and own the data develop a sensitivity to issues of data rectitude.



If data quality improvements are to continue to be the norm, the impetus must come both through technology architecture and business requirements. It must be understood and evangelized throughout the company that many times business continuity is data continuity: data problems invariably turn into business ones, increasing a firm's exposure to various unanticipated risks. In contrast to purely technical issues, a data steward will sometimes have to address and analyze data problems that are attributable to business routines that have a large human/manual element, and thus cannot be changed easily. The steward will need to have a business understanding and the wherewithal (support of senior management) to carry out the reengineering of various business processes. This may necessitate that they have a cursory understanding of various types of unstructured data (such as that found in electronic document management systems) and data that is not stored digitally, as well as the integration challenges associated with them. In such cases, a steward may have to help engineer a balance between the manual processes – routing paper requests, bar-coding, manual archiving – of a company's information center and records management databases and software.

Under a stewardship program, everybody shares responsibility and accountability for data excellence; information belongs to everybody, and like other (more empirical) corporate resources, it must be scrupulously managed and cared for. The stewards do not own the data; they define and bestow ownership and accountability accordingly, as they train, guide and mentor others in data quality best practices, rewarding compliance and adherence to quality specifications. Everybody that touches data throughout the organization must understand their role in data quality and be able to provide a feedback loop that will help stop bad data habits from propagating throughout the business. There will be pockets of resistance. However, a good lead data steward must know how to deal with the political side of the job. Drawing up a data consumer matrix, which describes who uses what data elements (and what actions are taken with that data), is always an invaluable exercise for assigning data trusteeship. This will help create a culture of data stewardship along all parts of the information life cycle and data supply chain – from the most important reference data to the most transitory of meta data.

In order to accomplish a firm foundation for continuous and measurable integrity of data, stewards should create and foster various partnerships and cross-functional teams (from business and technology) based on pre-approved agreements that govern systems not only inside, but also outside the aegis of enterprise control. Studies have shown that employee data entry errors are responsible for more than 75 percent of bad quality data. However, due to intensified e-commerce activity and broadening reliance on third-party data feeds, potential sources for substandard data are rising exponentially for many businesses. External third-party data sources that will commingle with (or feed) your internal system schemas should meet predefined standards of information quality and management before reaching the sanctity of your organization's gateway. Data suppliers must be held to a service level agreement (SLA) that states their responsibility and interest in providing quality data from a value-added information environment.

There is no one-size-fits-all job description for a data steward; stewardship roles will evolve according to the topology of IT infrastructure/architecture and the principal nature of the business. Care should be taken so that the steward is not saddled with more responsibility than he/she can handle. It will sometimes be good to have a hierarchy of steward roles or spheres of influence, where co-stewards oversee data probity issues along predefined systems or business segment domains in order to make data policy more manageable. Another possibility is to create a data stewardship committee that will share general responsibility for defining quality metrics; assessing acceptable and permissible use patterns of information assets; and assigning levels of stewardship responsibilities accordingly in order to limit data exposure and integration risks across all platforms. The committee will meet periodically to discuss all data governance issues, reporting on gaps in quality control and releasing/publicizing their findings. Items for committee consideration may include the following and beyond:

- Maintenance and storage strategies for corporate reference data.
- Comprehensive standards for data dictionaries, data models, calculation engines, business rules, ETL deltas and more.
- XML strategy: While XML and associated namespaces have been a boon for system-to-system connectivity and communication exchange, standardizing the syntax, format and relationships of information objects – a semantic consistency that reconciles business data and business-speak – will not always be easily achieved. People must know what XML tags mean in the real world business sense, across all of an enterprise's data structures, whether they are financial, manufacturing, etc. The stewardship committee will organize and establish correct linkages for all vital XML elements, artifacts and attributes via a meta data registry (MDR) that will document and define (i.e., capture meanings about) all XML namespaces by storing descriptions of data units and values across systems and data contexts.
- Data privacy and data security policies.
- Web-enabled universal "data shopping" on corporate intranet.
- Auditing of service level agreements to ensure proper data availability.
- Data standards: documented data domains, validations rules, data dependencies; source to target mappings; semantic management, naming conventions, data-typing, etc.
- Replication, backup and recovery policies and procedures.

With the relentless push toward federated-access systems that offer complex real-time (or near real-time) distribution and availability/exposure of data services, data volumes and information proliferation have grown to historically high levels. For many, data integrity problems have become worse than ever, establishing huge roadblocks for the timely delivery and sharing of quality data. An unceasing program of data governance and management, established by carefully selected people, roles, responsibilities and technologies, may be long overdue. Address your issues early (now!), not when you are ready to expose data to share. Address your enterprise-wide data quality issues with respect to accessibility, accuracy, consistency and completeness. Get executive buy-in and start your data stewardship program today!

William Laurent is the executive vice president of Loyer TCG where he leads the company's newly founded data warehousing practice. Laurent has a diverse systems background, successfully designing and managing the implementation of projects for the insurance, banking, finance, publishing, government, technology, entertainment and hospitality industries. The author of several white papers, he continues to be in demand as a data warehouse architect and lecturer. Laurent would enjoy receiving your comments, ideas or inquiries via e-mail at blaurent@loyertcg.com.