Zen and the Art of Information

Part II - Mastery in an Imperfect World

"Then, one day, after a shot, the Master made a deep bow and broke off the lesson. 'Just then, It shot!' he cried, as I stared at him bewildered." -- Eugen Herrigel

In the prior installment of Information Matters, I delivered the sad news that we will likely never be completely finished with our data quality work. The tools and processes will improve, but it may be decades before our systems are powerful enough to look at all business data from all possible angles and conclude it has perfect quality—let alone retain that perfection for the next hour, day or year. I argued that we need to be a bit philosophical about data quality and to settle in for a long journey. Akin to mowing your grass, it must be treated like a Zen practice: infinite.

For those of you who are now too depressed to put your fingers to the keyboard, read on. Though your work in data quality may never be completely finished, you can still achieve something profound and powerful for you and your business: mastery.

From Contention to Consensus – The Master List

In Eugen Herrigel's lovely book, *Zen in the Art of Archery*, the author takes up archery with a Zen master in post-war Japan in order to gain insight into Zen. Implicit in his search is the desire to achieve some type of understanding or mastery.

And so it is for data quality: there are large gains to be had if you don't try to do everything at once.

To begin, each company must determine what constitutes five-star data. For health care companies, it may be patient records. For manufacturers, it may be inventory or supply-chain information. For organizations in the public sector, it may be tax records or citizen information.

But whichever is the data of choice, chances are that numerous individuals, departments and applications must make decisions based on this data. Contention for this critical information introduces the potential for data quality problems, even as it signals broad consensus on which data really matters.

If only there was a high quality version of that information available at all times, to all applications—a Master List, if you will—then companies would gain tremendous insight and efficiency. This idea is at the heart of the remarkable rise in an emerging technology category called *data hubs*.

A data hub is a data quality solution that guarantees a master identity for a given business object, such as a customer or a product. At the core of a data hub is a centralized data

model. It is the rallying point to unify and reconcile common data across a collection of information systems. Using this data model, a data hub centralizes, de-duplicates, cleanses and enriches critical master lists of presumably critical data. The core data in the data hub is the minimal amount of data needed to guarantee uniqueness along with the references back to each source system. From there, anything can be added and shared.

As the name implies, data hubs are implemented using a hub and spoke model in which different applications subscribe to the data hub service. To date, most of the activity in this domain has centered on creating a master customer list, or *customer data hub.*

The Hub Spoke: "I, Customer"

According to Gartner, 75 percent of leading companies are incapable of creating a unified view of the customer. Why is this unification effort so difficult? Consider all the places that customer knowledge can be scattered throughout the enterprise. Your accounts receivable department knows who's bought something. Your sales application knows what they bought. Your customer service application knows which customers are having problems. And your web store knows their email address.

Each of these applications may have perfect data quality, but they have no way of reconciling all these dimensions of the collective customer identity. As a result, important decisions are made based on a faulty or incomplete view of who the customer is and what he wants. This has introduced a feeling of unease or even distrust of the data among top executives—not the happiest thought for an executive when she is staring at millions in capital costs and thousands upon thousands of IT staff-hours.

In many cases, this mistrust can be well founded. For example, in corporate governance, unclear revenue and profit reconciliation might expose corporate officers to serious legal risks. In finance, duplicate records can hide proper revenue recognition. In sales, an incomplete understanding of the latest customer-service interactions can hamper account reps. In order to resolve these omissions and discrepancies, you need a unified, reconciled, single source of high quality customer data.

Historically, there have been two primary approaches to creating this elusive Master Customer List:

- Synchronize data among key applications
- Collect information from all applications, clean it, and then place it in an data warehouse

In the first instance, each application owner establishes its own data management processes and resolves duplicate records. The more applications and business processes you have, the more complex the integration exercise becomes.

In the second option, you collect information from all your applications, clean it, and then place it in a central warehouse for reporting. Unfortunately, the cleansing process does

not push the clean data *back* into the transactional systems—so the root cause of the problem—the "here-ness and now-ness" of accurate data—is never really addressed.

Fortunately, there is a third option: store and manage key customer data centrally using a hub-based architecture. This results in less data movement, fewer integration projects, better data consistency, and standardized access to enterprise-wide information.

Data Hubs Oracle Style

Oracle data hubs are standalone services that enable companies to achieve high levels of data quality for their critical information assets. They establish centralized data definition and management without dependency on any specific application. As an open platform, Oracle data hubs unify operational, analytical, and legacy data sources.

The Oracle Customer Data Hub is the first of a broad series of data hubs currently in development at Oracle. Future hubs will include a *product data hub* that will help manufacturers and retailers to consolidate product information, a *citizen data hub* for public sector applications, and hubs for *financial consolidation* and *financial services accounting*.

Oracle's entry into the data hub market began well before it was even a product category. Actually, Oracle customers initiated the first deployments, as they unified their own separate CRM and ERP applications. For some customers, their home-grown customer data hub projects grew beyond the scope of their original intention and they called upon Oracle to help them craft a solution.

The central problem involved devising an effective, inclusive and future-proof customer data model. As it turned out, Oracle development had already created an integrated CRM and ERP customer model years before while laying the groundwork for the current generation of Oracle E-Business Suite. This has become a hugely important and valuable piece of intellectual property because it was developed and tested over several years in production environments.

As the word "hub" suggests, a data hub centrally services multiple applications with high quality data—creating the mystical *Single Source of Truth*. However, it is an important innovation that the hub does not have to *contain* all the data about a customer. Using a cross-reference system, the customer data hub is also able to deliver the vaunted '360 degree view'—that is, the ability to view every business transaction from a single place.

Long associated with data warehousing, the idea of a single source of truth has evolved in the data hub model to connote a "*living"* single source of truth. This implies continuity and highlights one of the key differences between a data hub and a data warehouse: a data hub is an active service, not a destination or a snapshot.

Of course, data warehouses and all manner of enterprise applications benefit from the fact that you do not *separately define* a customer master for the BI platform, the CRM system

and so on. Rather, these and other business applications access one common source of accurate data via the data hub.

Librarian with a Light Saber

Designing and deploying data hubs implies a positive decision to achieve mastery in data quality. But it does require significant time and effort. As Michelangelo said, "If people only knew how hard I work to gain my mastery, it wouldn't seem so wonderful at all."

The mere fact that there exists a title like Information Quality Professional indicates the importance of the task. Companies are spending big money on data quality initiatives because it is such a basic problem that influences so many aspects of the business. Who are my customers? What do they want? Do they have any money left? These are the kinds of questions that must be answered.

Finally, mastery implies that there is a master. Data hubs provide an invaluable information cross-reference, and their upkeep requires as much insight into the business as into the technology. Anyone willing to take this on rightfully deserves admiration and praise from their peers and gratitude from executives and stockholders.

In recognition of this, Oracle has bestowed upon these lucky professionals the title of *Data Librarian*. Honestly, some people don't exactly like this moniker, preferring to be called "data stewards" or "data custodians." Personally, I think data quality professionals should feel free to call themselves "Data Jedi," "Data Power Rangers," "Zen Data Masters," or anything they like. I won't tell.