RESULTWORKS

Transforming Strategy. Delivering Success.

KNOWLEDGE MANAGEMENT STARTS WITH HIGH QUALITY DATA

For many BioPharmaceutical companies knowledge management has been a keen interest for many years. Many programs have come and gone leaving a sour taste for knowledge management programs in general. Technology has often been the main focus of these initiatives. Significant energy has been expended in building data warehouses and seeking ever better search engines to pull together disparate data that is dispersed throughout the organization. The trouble is, in some organizations the state of the operational systems and the underlying data "takes a village" to piece together even routine searches for data. Instead of focusing on the top level knowledge management, greater effort needs to be spent on shoring up the foundational data and systems on which the business depends.

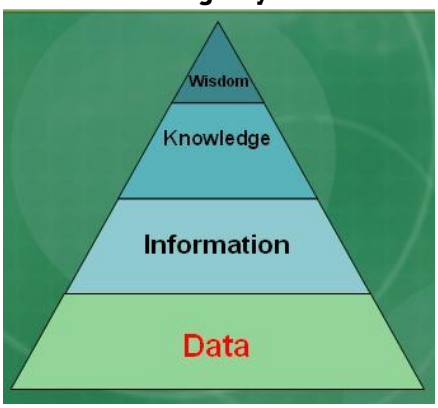
Data standards are often non-existent. Linkage of data and systems is dependent on people. The business is typically focused on the day-to-day efforts to move compounds through the pipeline in the most expeditious manner. This usually involves working around, not through, well-deployed systems with strong data governance. There is a cost associated with this patch-work approach to data management in terms of productivity, efficiency, throughput, overall quality, and even compliance.

How big of a problem is it? Often the business is so heads down that they do not have a good sense of data quality issues. In fact, they may think their data is just fine. An objective data quality assessment may be useful to illuminate problem areas and to develop an effective strategy for improving data quality.

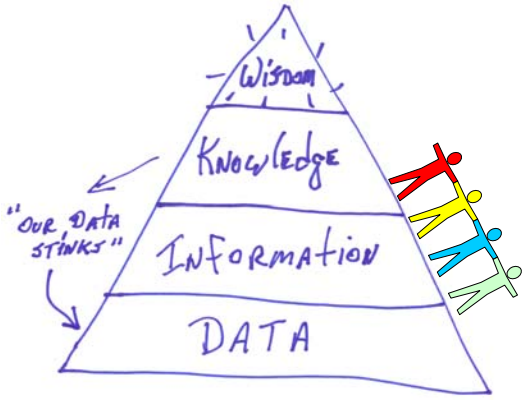
A five step approach for a data quality assessment includes:

1. Define the data
2. Determine data quality criteria
3. Analyze data in selected operational systems
4. Assess root cause of problems
5. Develop recommendations and a strategy forward.

Knowledge Pyramid



Reality Pyramid



Data Definition

At the risk of stating the obvious, data quality starts with the data. A solid definition of master data and key meta data is critical to the assessment. Master data is typically more controlled data that is used across a variety of departments and systems. Compound identifier is a good example of master data in life sciences. Meta data describes key attributes of master data and other data generated in the business. It is preferred that master data be controlled within one system and referenced from other systems. Automated references or links to master data is desirable, but often this data is manually entered into an operational system like LIMS (Laboratory Information Management System) or ELN (electronic lab notebooks) via research spreadsheets.

> There is a cost associated with patch-work data management in terms of productivity, throughout, quality, & compliance.

Once identified, it is also critical that the primary source of the data be identified. There should be one source of data that everyone references and that source should be controlled by the business. That source is where the data is accessed by researchers and other applications for security, performance, and other reasons.

Data Quality Criteria

Once the key data has been defined, the criteria for data quality must be determined. Criteria may already be established in SOPs, in data entry procedures and business rules, or in other data and system governance documents.

RESULTWORKS
Transforming Strategy. Delivering Success.

In some cases, determining the business rules around data entry may be the most difficult step. Quite often, the business rules are either inferred, assumed to be commonly understood, or left up to the discretion of the scientist. Relying on individual discipline to follow business rules is not likely to result in strong compliance. Without more system-oriented data entry checking, it is often difficult to establish and enforce business rules.

Analyze the Data

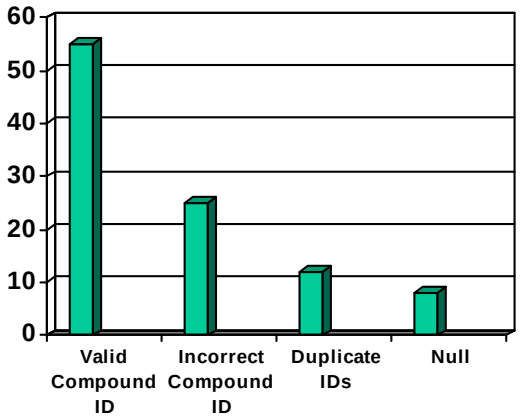
Once the data criteria are documented, analysis of the data in selected systems can readily proceed. Database tools can be used to construct, refine, and execute queries. Results can be captured and profiled to determine where issues exist and the severity of those issues. In some cases, profiling the data may allow for reverse engineering of business rules that are being followed as shown in the example chart below.

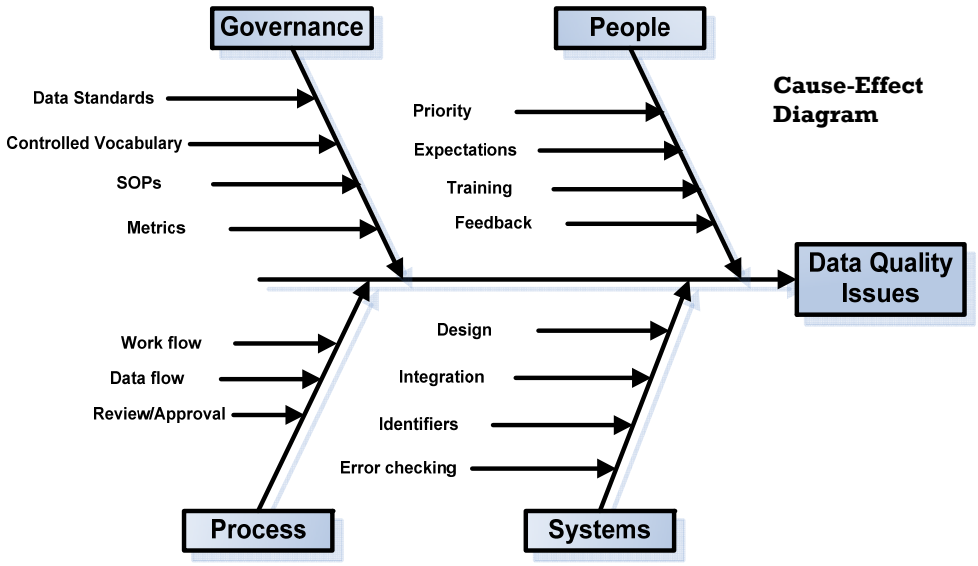
Assess Root Cause

After the data has been analyzed, problem areas standout. The root cause of those issues must be determined in order to correct the problems with the data as uncovered as well as the issue that is causing the problem in the first place.

As the cause-effect diagram shows, the root cause of problems in this area stem broadly from data governance, business process challenges, systems, and people or organizational issues.

Example chart of compound ID data quality





Recommendations

The root cause analysis highlights the most critical issues for the short-term. These deal with direct quality issues that need to be rectified immediately. Next, data standards and governance need to be addressed so that more formal business rules are applied, training can be conducted, and metrics can be put in place to set a baseline for performance. Longer-term, greater investments in automation can be made which will ultimately free people from spending inordinate amount of time in maintaining data quality manually.

Conclusions

High quality data is a requirement for an information strategy of R&D organizations. As a starting point, it may be useful to conduct a data quality assessment of key systems. The results of the assessment might be surprising, but they can be the driving force for business and informatics investments that will strengthen the foundation before making larger investments in a broader information management infrastructure.

Authored by and Bob O'Hara, Managing Partner and Co-Founder at ResultWorks, LLC based in King of Prussia, PA. For more dialogue on this topic, please contact Bob at bob.ohara@resultworksllc.com.