

DATA LIFECYCLE MANAGEMENT

USE CASE | Utility

Data Lifecycle Management

At a publicly traded gas utility company, a desire to implement Data Governance over critical asset data quickly evolved into a more strategic approach to do *data lifecycle management*.

Asset Analysis

A critical function performed by every gas utility company is studying asset condition, performance, and risk. The primary goal is to achieve public (and employee) safety and minimize risk. It is also important to have effective asset strategies to guide asset repair / replace decisions.

Data quality was a major barrier to significantly improving their asset performance and risk models.

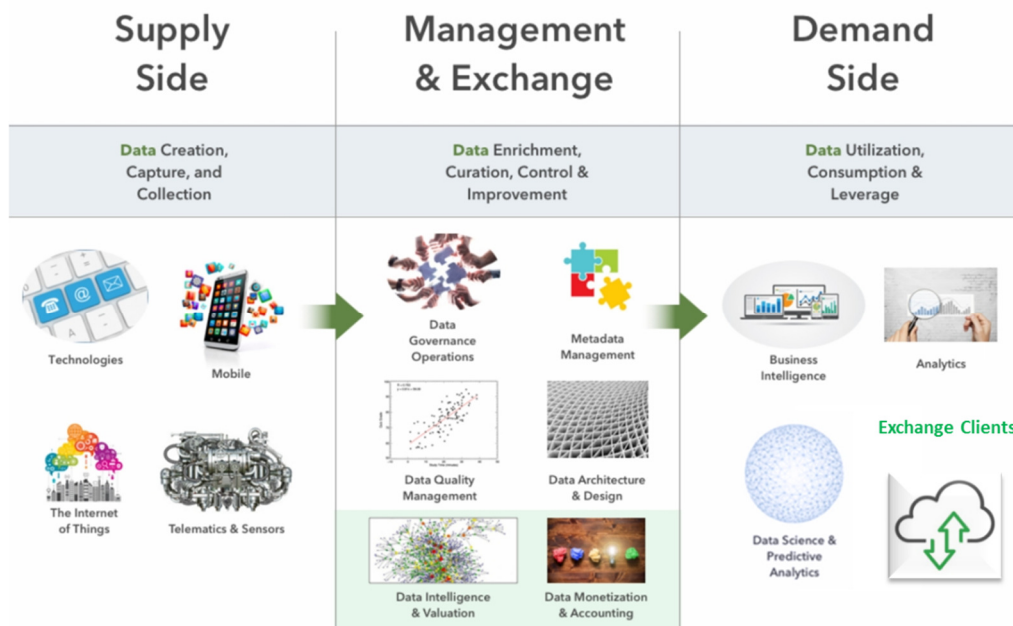
Working with the Corporate Analytics team for guidance on tools and advanced analytics environments, Integrity Management personnel were the early adopters to define the approach, workflows, and tools to manage the data lifecycle. The key steps of the data management lifecycle included:

- Data sourcing
- Metadata management and enrichment
- Data pipeline management and ETL capability
- Data quality profiling
- Business Rules development
- Data Quality Improvement
- Data curation for advanced analytics
- Use Case management

The model they implemented performs Data Governance by using planned Use Cases to identify value and prioritize efforts on which data to govern.

All data lifecycle management activities and supporting tools were provided by Infogix Corporation's product, Data3sixty. Data3sixty is a comprehensive toolset that operates as a Central Data Exchange (CDX) for data Suppliers, Managers, and Consumers.

Designed with an Amazon-like interface, analytics users can access the CDX to “Shop” and search for data, research its business definition, source and uses, evaluate quality scores, and then choose the “Ship To” function to deliver selected data to an output format of their choice, including a Hadoop advanced analytics environment.



Data Engineers can source and ingest data to be inventoried. They can also perform data transformations on the data “pipeline”.

Data Managers can govern the data coming into the CDX by inventorying the logical and physical metadata information and interrogating the full dataset values to quickly profile the data’s conformance to its defined format. This provides an initial data quality score.

Working with various business users, systems support personnel, and key stakeholders, Data Managers can elicit additional information about the data source, its business use, and relationships. This additional “conceptual” metadata can be captured to enrich the data’s definition and to define more advanced business rules and relationships that the data should conform to. The tool allows Data Managers to develop those business rules and process the data pipeline against them to more deeply assess data quality.

Tools are available for advanced users to cleanse, transform, and document the data pipeline so that it can improve and prepare data being used to support reports and other advanced analytics.

Workflows are designed with configurable triggers and tolerances to ensure data quality issues are researched and resolved. Remediated data can be mapped to a target format that can be used to reload an improved data set for the source system, if practical. This is helpful during systems replacements and data conversions.

The visibility to data quality scores and the enriched metadata from engaging business users serves as an effective continuous improvement loop to provide feedback to business end-users and IT support personnel to improve their business process and the User Interface (screens) used to collect the data from the data Supplier.

The CDX also allows Data Scientists to self-serve and populate data to research environments and to study methods that apply machine learning to improve the data. An aspiration for the Integrity Management group is to be able to develop models that can impute missing data values through machine learning in order to advance the state of analysis for assets. All of the information is maintained in a documentation repository to ensure any transformations or changes can be researched and explained.

Imagine a world where critical reports that support executive strategy decisions could state the information provided was “certified”, i.e. that you could state the data’s source, the data’s quality, and the quality of the process it came from.

Summary Key Points:

- Data governance has a negative and authoritarian connotation. Governance done poorly runs the risk of heading down a “boil the ocean” path that leads to high cost and little value.
- A Data Strategy that utilizes Use Case management to prioritize critical organizational data, and manages the data lifecycle from Suppliers through Consumers, is a more strategic way to improve data over time.
- Organize data lifecycle management around value, and set organizational goals to constantly improve critical data.
- Conclusion? Improving data quality is the key to growing an organization’s ability to leverage advanced analytics for competitive advantage.

Keller Schroeder's Data Strategy Group

At Keller Schroeder, we absolutely subscribe to the idea that Data Science, Machine Learning, and Artificial Intelligence are skills that every organization should have, and, in a connected, social media, *Internet of Everything* world, are vital to your company's future. **We think it is that simple - not easy, but simple.**

Keller Schroeder's **Data Strategy Framework** is a comprehensive implementation framework developed to help your organization establish the data lifecycle management practices necessary to successfully apply advanced analytics for your business benefit.

DATA STRATEGY
FRAMEWORK



Contact Us For Your
Complimentary
Data Strategy Framework

manage data. do science. get better

Prepared By

Tom Vargo
Director, Data Strategy Group
812-492-7341
tvargo@kellerschroeder.com

USE OF THIS INFORMATION: Keller Schroeder hereby authorizes you to view and distribute materials available on KellerSchroeder.com provided that the materials are used for internal informational purposes only. Any other unauthorized copying, alteration, distribution, transmission, performance, display, or other use of this material is strictly prohibited. Copyright © 2019 Keller Schroeder