Collibra

Customer Case

Using Collibra for A Data Quality Management Program at Aspen

Aspen Insurance is a leading insurance and reinsurance provider with a global footprint. Like many enterprises, Aspen has matured in terms of the quantity and diversity of data that it must manage. It has also experienced a shift from a process-centric to data-centric orientation as data has become increasingly shared and integrated.

Introduction

The increasing focus on data meant that data quality issues became increasingly important - and troubling. Data quality was less of a worry when Aspen consisted mainly of single transactional silos, but it simply had to be addressed once initiatives were undertaken to unlock value from data, such as building business intelligence environments. However, Aspen also realized that any data quality approach would have to include strong data governance to be successful. The company set about looking for a tool that could support this vision and selected Collibra as an enterprise-wide platform because it integrates data governance with metadata management capabilities needed for data quality.

"We realized that other companies had adopted a failing strategy by having a bottom-up approach to data quality where they measured a multitude of data quality points in isolation. These companies have ended up not knowing what they are measuring, where, why, who is involved, or what the consequences of a problem are. At Aspen we wanted a top-down data quality program from the start, and Collibra proved to the right platform for our needs." Aspen Data Quality Director..

The Data Quality Program

The approach Aspen took to data quality was not to view data elements in isolation, but to focus on data sets in the context of the overall data landscape. Additionally, Aspen wanted data quality to be an ongoing component of the operational environment - to be a program rather than a series of projects. These twin pillars - orientation to datasets, and operational integration - formed the basis of Aspen's data quality program.

Like all other financial services organizations, data is the lifeblood of Aspen and moves frequently from one data store to another. Each of the datasets that is in movement has its unique data quality governance profile that must be managed. This breaks down into understanding what the data in the dataset means, identifying individual stewards responsible for the dataset, and their roles, ensuring that appropriate data quality checks are in place, and planning for the mitigation of any data quality issues that might arise.

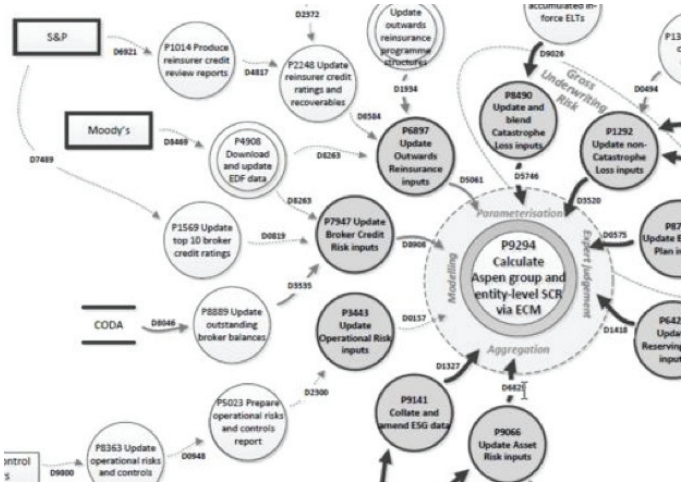

Figure 1: Visualization of Data Lineage for Data Quality Governance

Aspen began by defining data lineage across the data landscape. Metadata about datasets was entered into Collibra, from which it could be accessed in a number of ways, including the visualization shown in Figure 1. Each dataset was uniquely identified and given a profile that described what the dataset contained, in terms of business semantics. Most importantly, the stewards involved with each dataset from the point of view of data quality were identified using Collibra's RACI matrix manager (showing for each data quality task who is Responsible, Accountable, Consulted, and Informed). Figure 2 provides an example for one dataset.

One dataset typically has many data quality criteria, and each data quality criterion requires rigorous definition. Again, Collibra was able to provide the fabric in which to do this, and Aspen quickly worked out a metadata configuration for data quality criteria that was implemented in Collibra. Figure 3 illustrates one such criterion.

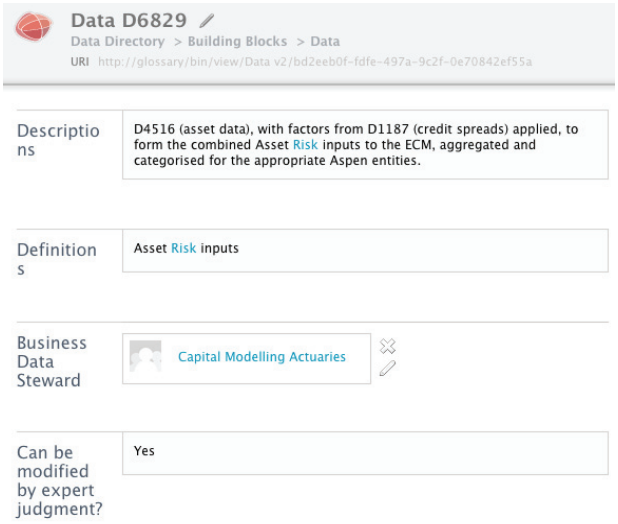

Figure 2: Dataset Data Quality Profile

As Figure 3 shows, the flexible nature of Collibra enabled Aspen to define precisely the metadata items they wanted for a data quality criterion. It was important for Aspen that Collibra did not force a fixed "one-size-fits-all" set of meta-data items on them for data quality governance. Further, Collibra's open, business-facing interface was ideal for data stewards in the business to review their involvement in the data quality program.

The next step from defining data quality criteria was to perform actual tests on the data itself. Details of how to test data quality points associated with individual criteria were also captured in Collibra, using its powerful business semantics features. These individual points were then implemented by developers in a SQL Server environment. Tests were run from this environment, and details of data quality exceptions collected.

SQL Server Reporting Services (SSRS) was then used to report back on these details, but the SSRS reports were also integrated with Collibra so that all appropriate metadata needed for each report was sourced from Collibra.

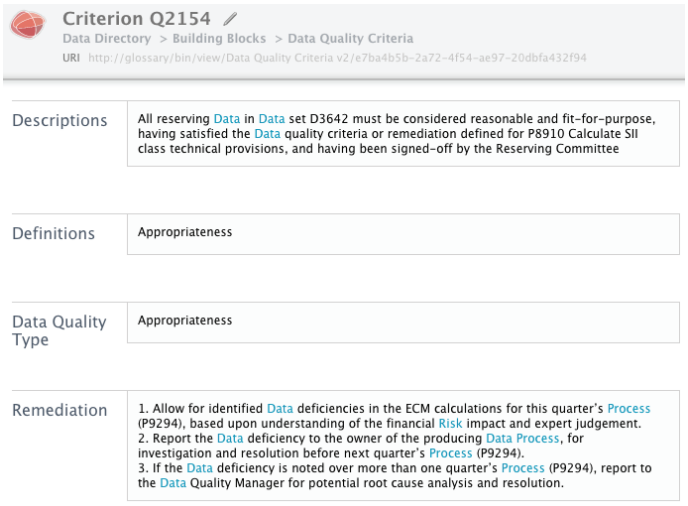

Figure 3: A Data Quality Criterion

The data quality reports are used internally, and also provided to external stakeholders, such as regulators to prove that Aspen has data quality under control.

Since Aspen's data landscape is constantly evolving, tackling data quality as a one-time project would have meant that it would have been out of date very soon after it was implemented. Instead, the collaborative features of Collibra and the governance framework it provides means that Aspen's investment in the data quality program is being enhanced over time, rather than being gradually eroded. Aspen has avoided turning data quality into a series of periodic, expensive, "clean-up" projects.

"From a technical data governance perspective, it was exciting to see how Aspen created a holistic data quality program, beginning with a single visualization of the data landscape, and going down to the individual data quality measures. The fact that Aspen was able to use Collibra for governance and metadata management of this program, and integrate Collibra with outside tools, is a testament

Figure 4: Example of a Data Quality Report