Data Quality at Kwantlen Polytechnic University

By David J. Greer

In the fall of 2010, Warren Stokes, Registrar for Kwantlen Polytechnic University (http://www.kwantlen.ca/) gave the presentation *Data Quality Assurance in the Registrar's Office* at the Canadian Banner User Group Conference in Victoria, BC. Stokes' presentation has many ideas that have broad application to improving data quality, process, and governance for any organization. This article shows measurable cost and administrative savings by implementing a proactive data quality process.

Background

Kwantlen Polytechnic University (Kwantlen) is a university located in British Columbia, Canada, with campuses in four locations: Surrey, Richmond, Langley, and Cloverdale. Kwantlen is primarily an undergraduate university with significant vocational (trades) training. In the fall of 2010 enrolment was just under 14,000 students across the four campuses. The Registrar's office has 110 staff members, including admissions, records, scheduling, and front counter.

Kwantlen's Enterprise Resource System (known as a Student Information System in the education market) is Banner, a third party application package used by many American and Canadian universities and colleges. Banner is used to track all student admission applications and the associated application approvals for admission to Kwantlen.

Challenges

In the fall of 2007, there were 7,400 new applications to Kwantlen. These applications created more than 17,000 internal records in the Banner system. Applicant and enrollment reports, critical to senior management, were taking 7-14 hours of staff time to prepare each and every week. Many of the data records had invalid data. For example, the city of Surrey was often spelled as Surrrey. When mail was sent to a student applicant with a misspelled city name, it was returned to the university as an invalid address.

Even worse the ranking system critical to admitting students at Kwantlen required redundant data to work. A free format field in each application record was used to record "special comments". These comments had to be, and were manually reviewed. Reporting staff were overwhelmed trying to clean up the data prior to report creation.

The key problem:

Data creators were not accountable for the data they were creating

There were a number of side effects of the redundant and incorrect data. Students were not being admitted to Kwantlen in a timely manner. If they were admitted, incorrect data would

prevent individual students from registering for the courses they wanted. Mail was returned on a regular basis which had to be dealt with on a case-by-case basis. Department moral was suffering due to the delays and mistakes.

Finding a Solution

The first part of the solution was to educate staff, fix systems, and improve data processes. The goal was to insure that an operator or data analyst could “tell the right story” two, five, or more years after the student had first been admitted to Kwantlen. The data in the Banner system should give a reasonable accounting of “what happened”.

Goals for the admission and enrolment reports were created. Key goals for these reports were:

- Extracted and published in less than 30 minutes
- 100% accurate at time of extraction
- No additional tuning (extract, transform, or load) prior to publication

Another task was to help the data creators become efficiently accountable for the data they were creating. To do this, a set of enrolment and application exception rules were created. These rules were used to create reports of invalid data on a daily or weekly basis and were automatically delivered to the data creators via email.

IT Aspects

The system Stokes and his team created at Kwantlen was a web based interface that allows users with the right permission to create, edit, and submit automated auditing rules. Each rule has a description, the database column headings to be included in the email message, the SQL statements to select invalid data, sort parameters, the frequency (daily or weekly), and an email address for delivery. For security reasons, only email addresses within the Kwantlen domain are accepted. Once the rule is verified it is submitted to an automated job scheduler (cron running on a flavor of UNIX in their case).

People, Process, and Change

Creating change in any organization is hard. Stokes and his team did a great job creating awareness of the value and of the need for change. Many IT systems and internal changes were made. Gaining acceptance of the changes among all the stakeholders required sustained effort.

When the daily data quality reports first started arriving they had many records to correct. Hundreds of records, in fact. It took time and resources for the various departments to spend the effort to start cleaning up the records. At first, it seemed like an insurmountable mountain to climb. Stokes encouraged all data owners to keep working on their reports. Initially, reports were run weekly, instead of daily. Over time, the data was cleaned up.

Old Habits Die Hard

After the initial clean up, data quality was good for a time. People go on holidays. Some change jobs or leave. New hires join the organization. After the initial success, Kwantlen experienced a return to significant invalid data, despite the daily audit checks. Some of the lessons learned were:

- Really, you do have to fix it every day.
- Bad data can happen even if you are on holidays.
- Distribution lists are a better idea than specific email addresses.
- A distribution list can have multiple people on it and is easy to change when someone goes on vacation, leaves, or someone new is hired.
- Old habits really do come back.

Birthdates and Addresses

Two data fields that commonly have invalid data at Kwantlen are birthdates and mailing addresses. It turns out that in Canada there are very few 13-year old kids who legitimately apply to university so doing a basic sanity check on the birthdate can avoid errors. Invalid mailing addresses cause all sorts of problems, as real paper documents must be sent to newly registered students at Kwantlen. To simplify address verification a third-party verification service is used.

Not all invalid data is created by Kwantlen staff. Students can apply on-line and it is often the case that students filling out on-line forms make mistakes. In the end, the data owners in the Registrar's office, must be responsible for insuring the accuracy of the data entered. The daily delivery of exception reports and daily cleanup by staff insure that the data is accurate.

Today

From hundreds of errors, there are now often no more than two or three in a given day. A total of sixteen daily auditing rules have been created and one weekly one. In 2010, there were 7,580 new students which caused 16,000 application records to be created. More students creating less work compared to the fall of 2007. Applicant and enrolment reports take 30 minutes to create compared to 7-14 hours previously.

Student applications are processed quickly. Applicants receive feedback faster. Key stakeholders are informed faster. With less work and effort for all involved. The data quality process saved time and money and helped improve the "customer experience" for the students.

This case study highlights the need for data management, data governance, and data quality. At MB Foster, we have helped numerous clients improve all three areas with solutions and experience that can make a measurable different to your organization.