USING THE DATA QUALITY SCORECARD AS A NEGOTIATION STRATEGY

(Practice-Oriented)

Traci Campbell
Acxiom Corporation
traci.campbell@acxiom.com

Keri Douglass
Acxiom Corporation
keri.douglass@acxiom.com

Wyndolyn Smith-Adams
Acxiom Corporation
wyndolyn.adams@acxiom.com

Abstract: How adversely would your product be affected by the loss or gain of a data source? Can a data quality assessment help your product's position in the negotiation? Through documenting a real-world case study, this presentation explores how one company's internal Data Quality Scorecard system was used as leverage at the negotiation table. The paper examines how data quality can drive service-level agreement standards, expose misconceptions about data dependencies, improve partnerships with vendors, necessitate internal process improvements, and ultimately increase a product's return on investment by driving costs down.

Key Words: Data Quality, Scorecard, TDQM, ROI, Negotiation Strategy

BACKGROUND

Throughout 2003 and 2004, Acxiom Corporation challenged itself to answer the question, "How's Your Data Quality?" As a result, this data integration organization innovated a corporate-wide data quality strategy called the Data Quality Scorecard (DQS); this unique system was based on Total Data Quality Management (TDQM) principles, implemented across multiple data products, and viewed as leverage for Acxiom's competitive advantage in the marketplace [1].

In the case study that follows, the authors examine how they were able to further the impact of the Data Quality Scorecard strategy by considering the return on investment (ROI) benefits of conducting such an analysis. In this real-world example, Acxiom exhibits a data dependency analysis (conducted using the Data Quality Scorecard approach) as a repeatable TDQM model used to assess data providers in a multi-sourced data repository; we identify where and when to take corrective action to improve data quality and prioritize the areas that can be a "deal maker" or "deal breaker" with data providers; and we discuss how a Data Quality Scorecard analysis can provide leverage at the negotiation table.

OVERVIEW OF THE PROBLEM

Our corporate-wide data quality strategy helps solve the challenge of measuring data quality for multiple information products within an organization. The DQS provides the framework for data products to consistently benchmark, measure, trend, analyze, and improve data quality at various touch points within a "data factory."

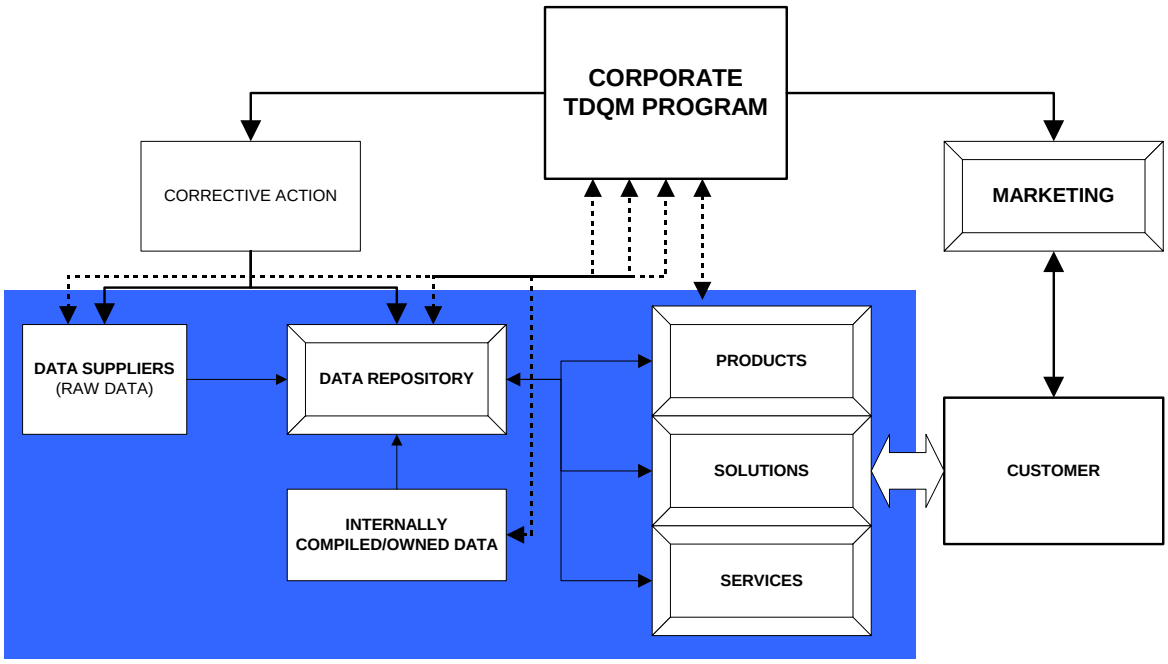

Figure 1: **Corporate Data Quality System.** This diagram represents the interaction of the TDQM system with the corporation's data factory. Solid lines represent the flow of communication and product quality requirements; dotted lines represent communication of and participation in the Corporate TDQM Program. Data quality measurements flow horizontally from "Data Suppliers" to "Data Repository" to "Product", which is where the focus of this case study lies.

The shaded area in Figure 1 represents the flow of data leading horizontally across the factory into data products. Raw data feeds into a "data repository" from which different data products extract what their customers need. If a particular data product is "multi-sourced", some of the raw data may be purchased from an external vendor and some may be compiled internally.

When our corporation first developed the DQS, we challenged ourselves to answer one simple question: "How's your Data Quality?". Once we gained control over our internal products, we began asking, "How does the quality of our data providers impact our product?", "How adversely would our products be affected by the loss or gain of a data source?", and "How can we use Data Quality in negotiations with our data suppliers?" This case study explores how we were able to answer those questions for one of our data products.

CASE STUDY: CONSUMER PHONE PRODUCT

Our organization builds a number of consumer data products, one of which is a telephone data product. For purposes of this study, we refer to this product as "Product A." Product A is a multi-sourced product — raw telephone data is compiled from internal sources, and raw data is purchased from external telephone data suppliers. Multiple data sources are combined to build Product A:

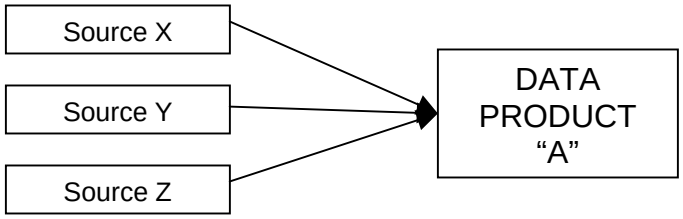

Figure 2: Multi-sourced Data Product "A".

Product A employs a corporate-wide DQS system and understands their data quality position in the marketplace. However, data sources are typically stable during assessment periods.

At contract renewal time, "Source X" decided to cancel their contract with Product A, meaning they would no longer supply data to the product. Source X was Product A's largest data contributor, the most costly, and presumably generated the highest quality data.

Analysis Approach

Product A's dependency on Source X was significant, but not fully quantified. Product A's company needed to assess the internal impact of losing their largest data provider. The product had high, perceived confidence in Source X and was convinced it needed to replace the source in order to maintain its data quality position in the marketplace and potentially improve ROI.

Without access to sample data from Source X, it was not feasible to conduct an internal data quality assessment to quantify the confidence. Even if data quality statistics from Source X were available, it would be difficult to quantify the impact on Product A after integrating additional sources. Our approach— rather than "quantify the loss" of Source X— was to evaluate the quality of Product A after integration *with and without* Source X and quantify the gap. This data quality gap could ultimately be used in negotiating a new contract with Source X.

In addition, to fully determine our organization's dependence on Source X, we knew we needed to examine the quality of "Product A without Source X" in the marketplace. While the DQS "meta-metrics" demonstrate how closely aligned Product A is to its data quality goals, Product A's essential concern was how the loss of Source X would affect its market position.

Data Dependency Analysis

In our data dependency analysis, our goal was to examine the quality of "Product A without Source X" in the marketplace. By establishing a competitive baseline, we could easily identify any accuracy and coverage gaps left by the loss of Source X as well as identify actions that would improve or maintain Product A's position given dynamic data sources.

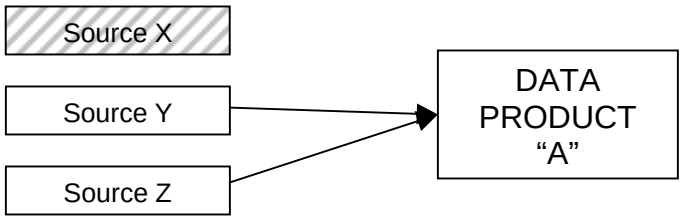

Figure 3: Data Product "A" without Source X.

The data dependency analysis was conducted in two phases:

Phase 1—

1. Assess Product A
2. Analyze competitive position

Phase 2—

3. Assess impact of changing data suppliers

1. Data Quality Scorecard Assessment (Phase 1)

Utilizing the DQS framework, we assessed Product A according to four key dimensions of data quality:

- Data Accuracy
- Data Completeness
- Data Access
- Data Consistency

By focusing on these key data quality parameters, we could eliminate or reveal areas such as customer relationship, price, and perceived value as driving factors in the negotiation.

2. Competitive Analysis (Phase 1)

Our organization then conducted a competitive analysis [2] to establish a marketplace baseline for Product A's data quality. By identifying key competitors in the industry, constructing a benchmark data sample, and retaining a 3rd party vendor to gather enhanced data from competitors, we were able to conduct a blind study.

Product A was treated as a "participant" in this study and was provided the same benchmark file to enhance as the other competitors. Product A enhanced the file through their normal fulfillment process, which included data from Source X.

All competitor files (including Product A) were analyzed using the Data Quality Scorecard framework and ranked accordingly.

	Product A	Competitor A	Competitor B	Competitor C
Data Accuracy	3rd	5th	2nd	**1st**
Data Completeness	3rd	2nd	**1st**	5th
Data Consistency	**1st**	4th	2nd	2nd
Data Access	4th	2nd	**1st**	3rd
Overall Score/Rank	2.79/ 2nd	2.57/ 4th	**3.35/ 1st**	2.77/ 3rd

Figure 4: Phase 1 Results—Using a Data Quality Scorecard, we measured how Product A's data quality *(with Source X included)* stacks up in the marketplace.

The results showed that Product A ranked 2nd out of four products (or companies), and varied in rank at the dimension-specific level. But the question still remained, "How will differing data sources affect Product A's position?"

3. Assess the Impact of Data Suppliers (Phase 2)

Using the same Data Quality Scorecard framework, we began to assess the impact that individual data suppliers had on Product A's data quality. The same benchmark sample from the previous competitive analysis was enhanced again for Product A, only this time omitting Source X from the enhancement process. This newly enhanced file was analyzed with the same data quality criteria used in Phase 1, and the results were applied to the Phase 1 Scorecard as if "Product A minus Source X" was a 5th competitor in the marketplace.

	Product A	Competitor A	Competitor B	Competitor C	Product A without Source X
Data Accuracy	3rd	5th	2nd	**1st**	3rd
Data Completeness	3rd	2nd	**1st**	5th	3rd
Data Consistency	**1st**	4th	2nd	2nd	N/A
Data Access	4th	2nd	**1st**	3rd	N/A

Figure 5: Phase 2 Results—Product A without Source X is treated as a 5th competitor in the overall data quality scorecard. Accuracy and Coverage were the key areas of concern.

Findings

The assumption going into the data dependency analysis was that Product A was dependent on Source X, and Product A would perform poorly in the marketplace without it. The organization presumed that Source X contributed the *most data* and the *most accurate* data to Product A. However, the Phase 2 results show that Product A performed the same in both coverage and accuracy *without* Source X.

Our findings in the Product A analysis proved that the highest priced data source did not necessarily command the highest quality data. By using the DQS, we were also able to show that Source X does not provide a competitive advantage for Product A.

Armed with this data dependency methodology for gap analysis, the organization can continue to examine the impact of Source X on *other* existing products within the organization to fully determine the dependency. This information can be leveraged in negotiations between the organization and Source X. And while accuracy and coverage were the primary considerations for Product A, other data quality areas (such as the customer relationship, price, reputation, and perceived value) can be discussed during negotiation and should be leveraged when pricing new or replacement data sources for Product A.

LEVERAGING THE RESULTS

In the case of the consumer telephone product, Product A, the data supplier prompted us to rethink the cost of our data. However, we soon discovered that cost was not the compelling factor for change. As Jack Olsen states in his book, *Data Quality: The Accuracy Dimension*, "Data quality concerns become additional justification for making a change initially justified by other drivers." [3]

The data dependency analysis revealed that Product A was "middle of the road" in the data quality arena, with or without Source X. There was room for Product A to improve. If the company agreed to cancel the Source X contract, Product A could use the cost savings to:

- evaluate non-traditional sources of replacement data
- improve processes that generate existing data
- assess additional data suppliers (vendors)

Data Acquisition teams could even analyze existing data suppliers using a similar process to the previous competitor analysis, and vendor "Scorecards" could be leveraged in contract negotiations.

Our Source Replacement Strategy

For our organization's consumer telephone product, Product A, it proved cost-effective to search for data suppliers to replace Source X data. With the DQS strategy in place, scorecards could be quickly produced and analyzed to see if the data quality criterion was met. Confirmed by Olsen, "Data purchased from outside the corporation… can be subjected to a quick data profiling process when received. Metrics can then be applied to generate a qualifying grade for the data source. It can help determine if you want to use the data at all. This can be used to negotiate with the vendor providing the data." [3]

The same four key dimensions of data quality used in the Product A, Phase 1 study were used to evaluate each replacement supplier. The replacement analysis methodology consisted of:

- Data Accuracy — name & address quality evaluation using internal data quality tools
- Data Completeness—data source samples were compared to existing internal files to determine amount of overlap
- Data Consistency — profile of data elements for validation against the vendor-supplied data dictionary
- Data Access/timeliness—recency of data was determined by matching to existing, internal data files

As in the case study, each replacement supplier scorecard was rolled up into a "master" scorecard for comparison and data quality rankings. This Master Data Quality Scorecard provided the Data Acquisition team a complete view of the potential replacement suppliers' quality.

	Source A	Source B	Source C	Source D
Data Accuracy	3rd	4th	2nd	**1st**
Data Completeness	3rd	2nd	**1st**	4th
Data Consistency	**1st**	4th	2nd	2nd
Data Access	4th	2nd	**1st**	3rd
Overall Score/Rank	2.79 / 2nd	2.57 / 4th	**3.35 / 1st**	2.77 / 3rd

Figure 6: Replacement Data Supplier Master Scorecard—new sources are evaluated, ranked and compared prior to making a data purchasing decision.

The Master Data Quality Scorecard provided the business intelligence that could drive negotiations with replacement vendors. Although rankings were established, no vendor was a clear standout in terms of actual quality numbers—the range of scores and distribution of dimension-level rankings was quite close. This triggered cost to become a negotiation consideration.

Because of good vendor relationships, and given the closeness of the data quality results, our organization was able to obtain price concessions and/or additional data for evaluation from the replacement suppliers. "Price to acquire the data" was added to the Master Scorecard, and the cost and data quality of keeping Source X was displayed for consideration.

	Source A	Source B	Source C	Source D	Product A (with Source X)
Data Accuracy	3rd	4th	2nd	**1st**	3rd
Data Completeness	2nd	3rd	**1st**	4th	3rd
Data Consistency	**1st**	4th	2nd	2nd	**1st**
Data Access	4th	2nd	**1st**	3rd	4th
Overall G.P.A./Grade	3.00 / 2nd	2.57 / 4th	**3.35 / 1st**	2.77 / 3rd	2.79/ 2nd
Cost	$1 mm	$500K	**$1mm**	$750k	$2.5mm

Figure 7: Data Quality/Cost Master Scorecard—new sources are evaluated, ranked and compared according to data quality, and price to acquire the data is considered prior to making a data purchasing decision.

Equipped with both price and quality figures, Product A was able to drive negotiations with both Source X and potential replacement suppliers to maximize data quality and cost savings.

Examining the Data Quality/Cost Master Scorecard reveals that "Source A" was second place overall in data quality, but the same acquisition price as the first place/highest quality replacement "Source C." Product A leveraged this information to negotiate a price reduction from Source A. The original cost of Source A was reduced from $1 mm to $800,000. The net result was that an additional $200,000 was saved without significantly comprising data quality.

Risks for Source Replacement or Process Improvement

There are risks associated with any strategy to leverage data dependency results. First, other products within the enterprise besides Product A may be dependent upon Source X. They should be brought into the negotiation process as a secondary data user, and a data dependency analysis should be conducted for each product prior to negotiation.

Secondly, all products involved throughout the organization must evaluate the expense and resource costs associated with build processes that eliminate Source X. Products must ask themselves, "How easily and quickly can our process be redesigned to incorporate a new source?"

Lastly, very often in the data industry, the data source being evaluated may also be a competitor. While partnering can be a lucrative business decision, companies must weigh the risk of using competitors' data in their products prior to negotiating with the data suppliers.

USING DATA QUALITY TO DRIVE NEGOTIATIONS AND SET PRICING

The cost savings realized from the cancellation of Source X was so significant, it could be used to replace the external data source as well as improve efficiencies in data quality internally. This type of analysis and corrective action enabled Product A to improve overall in the marketplace in both data quality position and price offering to its customers.

The data dependency analysis exhibited in this case study is a good example of how the DQS system can help drive standards for Service Level Agreements (SLAs) and data supplier contracts. This will help to improve market position. Once DQS baseline measurements are established, future sources must adhere to or exceed that level as the "norm", thus ensuring the quality of the data product. Product negotiators can even stipulate consequences for existing suppliers that fail to meet quality baselines — such as requesting a cost reduction, seeking a monetary refund, or even canceling the contract altogether

Including DQS measurements in SLAs with data providers will ultimately benefit the organization's own data product. However, in order to ensure the SLA is met, product build teams must maintain a data quality measurement process. Poor or non-existent data quality monitoring processes can force companies into situations where data products are relying on too costly data suppliers with little to no real value-add. If companies can obtain high quality data sources at lower costs, or improve internal process efficiencies, they can pass that savings on to their customers when price setting.

Scorecard results may also expose areas where suppliers could improve the data they are providing. For example, in this case study, individual scorecards were shared with the sources evaluated so suppliers were aware of their data quality performance. With this information on hand, negotiations could thereby include a partnership on data or services. For example, if a data supplier exhibits poor address quality, and your company can help improve that, negotiations might include partnering your services: "Our company can provide address cleansing services in exchange for your phone number data." Partnership opportunities can further drive costs downward while simultaneously improving vendor relationships.

CONCLUSION

Our organization continuously applies the Data Quality Scorecard framework to products, data sources, and product delivery processes. These scorecards are beneficial to the data supplier negotiation process, pricing models, and contract renewal terms.

Acxiom was able to assess data providers in a multi-sourced data repository through a data dependency analysis based on the repeatable DQS system. The organization identified where and when to take corrective action to improve data quality. The DQS was able to provide leverage at the negotiation table by prioritizing areas that were either successful or ineffective in the negotiation process.

By taking the DQS system a step further from "How's your Data Quality?", our organization is now able to answer, "How does the quality of our data providers impact our product?", "How adversely would our products be affected by the loss or gain of a data source?", and "How can we use Data Quality in negotiations with our data suppliers?" By specifying data quality requirements in the SLA, monitoring data quality through the DQS system, and identifying partnership opportunities for data quality improvement, other data companies can ensure healthy vendor relationships and successful negotiations with data suppliers.

REFERENCES

[1] Campbell, Traci and Wilhoit, Zachary. *How's Your Data Quality? A Case Study in Corporate Data Quality Strategy.* International Conference on Information Quality, MIT, 2003.

[2] Smith-Adams, Wyndolyn and Talburt, John R. *Conducting an Information Product Competitor Analysis: Case Study.* International Conference on Information Quality, MIT, 2003.

[3] Olsen, Jack E. *Data Quality: The Accuracy Dimension.* Morgan Kaufmann, © 2003.