“Re-Engineering the US Contact Center Customer Data”

1-Jul-03

Delphine Clément, Pascal Vouille
Customer Knowledge Management, Data Quality Branch
Hewlett-Packard Company

Abstract

Based on the IQ Principles and Foundations taught during the MIT IQM I session, we plan to build a solid and useful process to identify, understand and solve the current data quality issues faced by the HP US Contact Center.

Our contribution to the IQM I methodology will be to demonstrate that the concepts like data quality dimensions, 3 C’s, data quality assessment /measurement process, root-cause analysis, IP Maps, not only apply but also bring an objective value-added to a real business case.

Below is an outline of the proposed project phases.

Phase I: IDENTIFY

Data Flow Map

The aim of this map is first to document the information or data flow and processes. This step will also help us identify the Collector/Custodian/Consumer interactions with the data. Finally, we plan during this exercise to understand where the data is transformed and where it makes the more sense to measure the data quality.

Objective and Subjective Data Quality Measures

The objective data quality dimensions that we will measure are the: completeness, timeliness (age of data only), customer address validity, accuracy, uniqueness and amount of information, provided the Data Consumers tell us against which “reference database” this last measure should be conducted.
To perform the objective measures, we will develop ad-hoc queries and run them in the USCC production database.
Nonetheless, accuracy will be measured through outbound calls to a random sample of customers.
Address validity will be measured by running the data through an address validation tool, be it Firstlogic or Address Doctor.

The subjective data quality dimensions that we will measure are fourteen in total: completeness, timeliness, consistent representation (HP terminology is standardization), accuracy, uniqueness, objectivity, believability, value-added/relevancy, proper amount of

information, ease of understanding/interpretation, accessibility, security, reputation, ease of manipulation.
We voluntarily re-grouped relevancy with value-added, and interpretability with understandability (compared to the initial sixteen dimensions of the IQM I Principles and Foundations) as, after multiple thorough discussions, we agreed that for HP, the notions were similar.
To perform the subjective measures, we have adopted the survey tool process and therefore internally developed a web-based questionnaire. The questionnaire will be made of four sections (Section 1: survey-taker info, Section 2: dimensional data quality assessment, Section 3: dimensional data quality importance rating, Section 4: business-related open questions) and will be sent to a 10% random sample of US CC data collectors, custodians and consumers.

Gap Analysis
We will propose different types of gap analysis as an outcome of the measurement step detailed above.
For example:
Vertical gap: analysis of the discrepancy, within one C boundary (collector, custodian or consumer), between the data quality dimensions **importance rating** and the data quality **perceptions**.
Horizontal gaps:

- Across the 3 C's boundaries, analysis of the discrepancy in terms of data quality dimensions **importance rating** between Collectors, Custodians and Consumers of the data.
- Across the 3 C's boundaries, analysis of the discrepancy in terms of data quality **perceptions** between Collectors, Custodians and Consumers of the data.

We envision that the horizontal gaps will reveal more root-causes of the data quality issues (communication issues, different set of expectations, etc…) than data quality issues themselves.
Objective/Perception: analysis of the discrepancy between the objective measures performed against the database and the perception measures, for the same set of data quality dimensions.

Cost of Non-Quality Analysis
With the results of the measures and of the gap analysis, we will try to evaluate the cost of the Non-Quality for this particular case study.
The evaluation will be both qualitative (impact on the business efficiency) and quantitative (dollar amount).

Phase II: UNDERSTAND

Root-Cause Analysis
At the end of phase I, we will be ready for the root-cause analysis: the data quality awareness will have been created, the measures and the gap analysis will have helped define the data quality issues.

We will use one or more of the graphs presented during IQM I (fishbone diagram for example) to locate the source of the data quality issues.
As of now, we know that this phase will require discussions with and inputs from the 3 C's as well as the USCC management.
Root-Cause analysis is the only way to ensure a sustainable data quality and data quality program.

Phase III: SOLVE

Solution Recommendations
With the information gathered during phase I and phase II, we will have gained sufficient knowledge to recommend appropriate and value-added solutions to the USCC business. We today think solutions may be multiple, going from process re-engineering, infrastructure modifications, change management, training, communication and budget allocation.
IP Map: when elaborating the solution recommendations, we will re-use the initial data flow map and transform it into an IP Map. Indeed, we believe that punctual data quality checks will be necessary throughout the information/data manufacturing cycle (from data collection to data consumption) in order to guarantee that the data product is compliant with the expectations/specifications.
Monitoring: when proposing solution(s), we will also recommend to put in place monitoring processes/tools. We investigate the possibility to adapt the database schema in order to capture the key quality indicators and also the possibility to use control flow charts as mentioned during IQM I.
We will not be the ones to actually implement the solutions, as we are simply a data quality consulting body for the HP businesses. We will however make ourselves available for the USCC business during the implementation phase.

Project Timelines
We plan to complete the entire loop by early January 2004 at the latest. However, this is just a guesstimate as we might face unexpected and various delays.
Today, we have already completed partially step I of phase I and have also partially designed the survey questionnaire.