

# Rapid Data Quality Assessment

## Rapid Data Quality Assessment: An Introduction

Vague feelings of poor data quality can hamper business activity by casting doubt on the data that feeds both operational and analytical processes. Yet the absence of clearly defined measurements to demonstrate how business impacts are attributable to erred data prevents developing an appropriate business case for introducing data quality management improvement. When resources are allocated to correcting “bad data” without being able to evaluate the root causes there is no ability to objectively evaluate the relationship between poor data quality and business performance. This suggests a need for an approach that can quickly identify high priority data issues whose remediation can be justified.

The objective of Knowledge Integrity’s rapid data quality assessment is to meet this need. Relying on our many years of experience in data quality analysis techniques, Knowledge Integrity analysts can execute this assessment as well as provide training and knowledge transfer, with a rapid schedule that can be adapted to best suit the client’s needs. The resulting report will describe and prioritize clearly identified data quality issues, recommendations for remediation, and suggestions for instituting data quality inspection and control for data quality monitoring. In addition, our analysts will provide templates and train your team members so that they can perform the rapid data quality assessment on subsequent data sets.

## Objectives

Through proper planning and the right expertise, Knowledge Integrity analysts can work with the client to perform the rapid data quality assessment within a 3-4 week period. Our experienced analysts using our planned methods will help identify high visibility data issues as well as characterize the business impacts incurred by those issues. At the same time, opportunities for improvement can be identified, providing an ability to provide an objective assessment of critical data, determine whether the levels of data quality are sufficient to meet business expectations, and if not, evaluate the value proposition and feasibility of data quality improvement.

*During this process, the KII team will identify immediate opportunities for improvement using a combination of careful selection of data for focus based on perceived business impacts and potential data issues. The process consists of these phases:*

- 1) *Planning*
  - *Assessing the scope*
  - *Reviewing the business objectives and critical data sets*
- 2) *Business process evaluation*
  - *Business impact analysis*
  - *Business process information flow mapping*
  - *Evaluation of system documentation*
- 3) *Preparation and Data Analysis*
  - *Isolation of critical data elements*
  - *Definition of data quality measurements*
  - *Prepare data analysis environment*
  - *Analysis to capture measurements*
- 4) *Synthesis of analysis results*
  - *Review analysis results*
  - *Enumeration of potential anomalies*
  - *Potential drill-down for additional detail*
  - *Preparation of assessment report*
- 5) *Review with business owner*
  - *Present assessment report and associated measurements*
  - *Prioritize discovered issues*
  - *Identify remediation tasks*
- 6) *Training and knowledge transfer*
  - *Training session*
  - *Provision of templates*

## Planning

At this stage, our data quality assessment plan template is adapted to the particular environment. Client team members are identified, internal preparation tasks are assigned, and the schedule is aligned to best meet the client's needs.

## Business Process Evaluation

Since the process is directed at linking data quality issues to business impacts, it is necessary to select a specific business process that is impacted by poor data quality. During this step, the analyst maps the information flow supporting the business process, identifying where data is created, exchanged, or used, and selecting key locations in the process flow where there are critical dependencies on data that would be appropriate locations to probe the data. Once business process, data set, and location in the process have been selected, the KII analyst will consult with client subject matter experts to identify specific business impacts associated with data errors and determine if it is possible to quantify (at a gross level) the cost value of those business impacts.

## Preparation and Data Analysis

The typical process of data profiling is undirected, allowing the tool to drive the activity, with the analyst drilling through various measurements with little forethought, seeking any potential anomalies that may indicate a data flaw. Alternatively, in this process, the next step is to assess the quality of the data in the context scoped by the results of the business process evaluation. This means that the focus is limited to evaluating the quality of the data elements potentially implicated in the identified business impacts based on the expectations of the business users. This incorporates these activities:

- Isolation of critical data elements – reducing the scope of the data elements to be examined to those that are critical to producing the results that are impacted;
- Definition of data quality measurements – reviewing the issues that are reported by the business users to provide specific types of data quality expectations, what is to be measured, and how those measurements reflect the business impacts;
- Prepare data analysis environment – At the same time, the client will prepare data for analysis, either through the use of their own tools or using KII-supplied methods;
- Analysis to capture measurements – analyzing the data and then capture analysis statistics and measures.

## Synthesis of Analysis Results

Even with constraining profiling to those critical data elements, any conclusions regarding the quality of data within the specific business context require review of the analysis results synthesized into a coherent enumeration of potential anomalies, annotated with descriptions indicating their potential relevance. This phase includes these steps:

- Review analysis results – examining the results of the analysis statistics within the context of the business impacts;

- Enumeration of potential anomalies – specifically describing the potential anomalies along with the notes explaining why those anomalies are related to business impacts;
- Potential drill-down for additional detail – some discovered items may require additional drill-down or review, perhaps using other analytic tools;
- Preparation of assessment report – this provides a tangible artifact to the client detailing the potential data flaws and also provides a template for iterating the process internally on another data set.

## **Review with Business Client**

The KII consultant will meet with the client to review the discovered anomalies and explore options for remediation. This phase includes these steps:

- Present analysis report and associated measurements – walk through the anomalies, explain what measurements were performed, the scores, and why those scores may be related to identified business impacts;
- Prioritize discovered issues – based on the business client input, prioritize the issues based on significance, business relevance, and feasibility of remediation;
- Identify remediation tasks – list specific tasks to be performed to correct data, evaluate root causes, mitigate immediate issues and remediation steps to eliminate the root causes of the introduction of flawed data.

## **Training and Knowledge Transfer**

A training session will review the process with selected staff members to be tasked with repeating the process. In addition, as part of the knowledge transfer, KII will provide templates for the repeatable process, including a task plan, capturing profiling statistics, and the profiling report template, along with training session material.

## **Rapid Data Assessment – Tangible Results**

This process would not take the place of a full-scale data quality management program, but will establish the context that can be used to establish the business case for investing in data quality improvement. Alternatively, the process can be used to evaluate whether the costs associated with data quality improvement would provide a reasonable return on the investment. Either way, this repeatable assessment process provides tangible results that can either validate or negate the “fuzzy” organizational perception of poor data quality, and provide the impetus to make the organizational commitment to measuring, monitoring, and improving data quality.

## **For More Information and Pricing**

For additional details, and pricing quote, contact:

David Loshin

[loshin@knowledge-integrity.com](mailto:loshin@knowledge-integrity.com)

301-754-6350