

## Results

---

The objective of the effort is to develop a tool to support the identification of data quality issues and the selection of tools for addressing those issues. In order to determine what features would be needed, the following questions were initially developed to be asked of the data:

- Is your data complete and valid?
- Does your data contain fields that must be split into smaller parts before entering the data warehouse?
- Does your data have abbreviations that should be changed to insure consistency throughout the data warehouse?
- Is your data correct?
- Is there redundancy in your data (is the same information in more than one place in the various databases that you are drawing from)?
- Do different forms of data need to be converted to a single form for consistency across the data warehouse?
- How well does the data reflect the business rules? Do you have missing values, illegal values, inconsistent values, or invalid relationships?
- Do you have free form text that needs to be indexed and classified to be useful in the data warehouse?

This list of questions was reviewed by members of four New York State agencies in the initial stages of developing the framework for data repositories within the agencies. Based on their review, additional questions were added.

A matrix (Table 1) was developed that mapped the features of the data quality tools to the questions that were asked. Examples of tools that contain the features are also part of the matrix. The matrix was reviewed by IT professionals from four New York State agencies. Based on their review, additional questions were added to the matrix. This matrix can be used by builders of data warehouses in the initial stages of development to evaluate their data sources. Once the questions have been asked of the data, the warehouse developer will be able to identify problems in the data sources. The data quality tools have different features to address specific problems in the data. The "Mapping Data Problems to Features of Data Quality Tools" matrix in Table 1 will allow the warehouse developer to focus on which features are needed to address specific problems in the data sources. For example, if the data sources contain primarily name and address data, then a data cleansing tool may be sufficient. On the other hand, if most of the data is financial, then an auditing tool may be more appropriate.

Table 2 contains information about specific tools, including URL's, price, platform, and special features of the tool. The matrix can be used to begin evaluation of specific tools.

### Table 1- Mapping Data Problems to Features of Data Quality Tools

[View larger image of table.](#)

### Table 2 - Data Quality Products

[View larger image of table.](#)