

Data Profiling

Data profiling is the process of examining the data available in an existing data source (e.g. a database or a file) and collecting statistics and information about that data.

The purpose of these statistics may be to:

- Find out whether existing data can easily be used for other purposes
- Improve the ability to search the data by tagging it with keywords, descriptions, or assigning it to a category
- Give metrics on data quality, including whether the data conforms to particular standards or patterns
- Assess the risk involved in integrating data for new applications, including the challenges of merged data sets.
- Assess whether metadata accurately describes the actual values in the source database
- Understanding data challenges early in any data intensive project, so that late project surprises are avoided. Finding data problems late in the project can lead to delays and cost overruns.
- Have an enterprise view of all data, for uses such as master data management where key data is needed, or data governance for improving data quality.