**Data Profiling**

**Data profiling** is the process of examining the data available in an existing data source (e.g. a [database](http://en.wikipedia.org/wiki/Database) or a [file](http://en.wikipedia.org/wiki/Computer_file)) and collecting [statistics](http://en.wikipedia.org/wiki/Descriptive_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](http://en.wikipedia.org/wiki/Tag_(metadata)) it with [keywords](http://en.wikipedia.org/wiki/Keywords), descriptions, or assigning it to a category
* Give [metrics](http://en.wikipedia.org/wiki/Software_metric) on [data quality](http://en.wikipedia.org/wiki/Data_quality), including whether the data conforms to particular standards or patterns
* Assess the risk involved in [integrating data](http://en.wikipedia.org/wiki/Data_integration) for new applications, including the challenges of [joins](http://en.wikipedia.org/wiki/Join)
* Assess whether [metadata](http://en.wikipedia.org/wiki/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](http://en.wikipedia.org/wiki/Master_data_management" \o "Master data management) where key data is needed, or [data governance](http://en.wikipedia.org/wiki/Data_governance) for improving data quality.