

**IQ Tool Link: Thomas Frisendal:  
Free tool for automation of SQL Server repeatable data profiling**

The easy way to load Microsoft SQL Server data profiles into a database! Automate your profiling activities by way of building a data quality repository database in e.g. SQL Server.

By doing that you can repeat the profiling tasks and you can compare results over time. This is the obvious way to work seriously with data quality on a repeated basis - enabling you to answer questions like "Is the quality of our Product data increasing over time?"

The current version of the Data Profile Task in SQL Server 2008 does indeed provide you with most of what the need in terms of data quality measurements. The profiling results are kept in XML files, which are nicely structured according to a quite complex schema.

However, as a result of that, it is not straightforward to load the results into a relational database. The challenge is the deep hierarchical structure of the XML files.

Using XSLT style sheets on Data Profiling XML files I have made a solution to this problem, which I offer you free of charge (and un-supported). Using XSLT style sheets you can reformat the XML files produced by the Data Profile Task. The resulting, reformatted, files are loaded 1:1 into simple tables, one for each type of profile:

- ✓ Candidate Keys
- ✓ Column Null Ratios
- ✓ Column Patterns
- ✓ Column Statistics
- ✓ Functional Dependencies
- ✓ Inclusions (between tables)
- ✓ Lengths of column content
- ✓ Value distributions of column content

The loading can easily be accomplished using normal SSIS data flows, which you must set up yourself. The table design of the data profiling repository database is also left to you to decide.

<http://www.informationqualitysolutions.com/page2/page11/page13/page13.html>