



Vincent Wylenzek

Vincent is a senior Business Intelligence architect with broad experience within the business intelligence domain. Over the last 8 years Vincent worked successfully on BI projects for more than 20 clients in a wide variety of roles. Vincent has hands-on experience with the big-five ETL and reporting tools and is certified in the Microsoft BI stack. Besides his deep technical knowledge, Vincent is educated in the finance and accountancy & control domain. Innovation, efficiency and personal improvement, ambition, drive to learn new things are key.

A red handwritten signature, likely of Vincent Wylenzek, written in a cursive style.

"How did I get this figure? This can't be right - where did it come from?"

We are regularly asked these questions. Evidently, the figure reported is not the one that was expected. The correctness of the figure is determined by the quality of the source data and all the transformations and manipulations that took place on that figure.

More and more often, clients want to be provided with insight into data flows, the operations and the business rules applied. This desire is fed by legislation, for example, Sarbanes-Oxley for (American) listed companies, the Dutch Corporate Governance Code, Solvency and BASEL, and by the increased risk-awareness of entrepreneurs.

Various ETL tool suppliers have received requests from clients and are therefore developing data lineage tooling, or have already released a version. This tooling allows end users to request information about the origin of the data. However, the ETL tool suppliers' data lineage tooling has its limitations. Apart from vendor-lock-in, this often involves the operations that take place within the tool. Information of operations within the reporting environment are out of scope (cube, semantic layer, reports and the like). This problem is less severe if the tool supplier supplies the complete BI stack as well as the ETL tool. Microsoft supplies the complete BI stack, but has not yet progressed far enough with its data lineage tooling.

How can this demand for information be met without the data lineage of the tool supplier?

To this end, Ordina developed a custom data lineage solution in a client project. We worked with a data vault at the client's place of business. On top of this, we used various dimensional data marts, and the complete Microsoft BI stack for the tooling. Our custom data lineage solution has a central metadata repository. All the business rules are stored in this repository, and they can be modified via a GUI (like Microsoft MDS). The specifications of the rules are stored in the repository and they can be reported on by the reporting tool (SSRS).

In addition to setting the business rules, it is also necessary to keep the ETL mapping of all entities and attributes updated. In this way, for all attributes in every layer within the architecture, it is clear how the figure is being processed (see figure 1).

If all of this has to be entered manually, the error-sensitivity factor is high. For that reason, we use a custom-made tool with which we can read out the definition of all the SSIS data sources used (mainly OLEDB). However, if the query string is built based on one or, underwater, more variables, we then enter the data manually. For the ETL we make use of several standard templates for which we restrict the number of variables. When the choice is made to generate ETL, the query string variables can actually be completely eliminated.

The biggest advantage of this custom data lineage approach is that the client itself has insight into which transformations and calculations have taken place on a particular figure. If the client still believes that the figure is not correct, the reason can be found more quickly and categorized as a data quality problem or an incorrectly applied or wrong business rule.