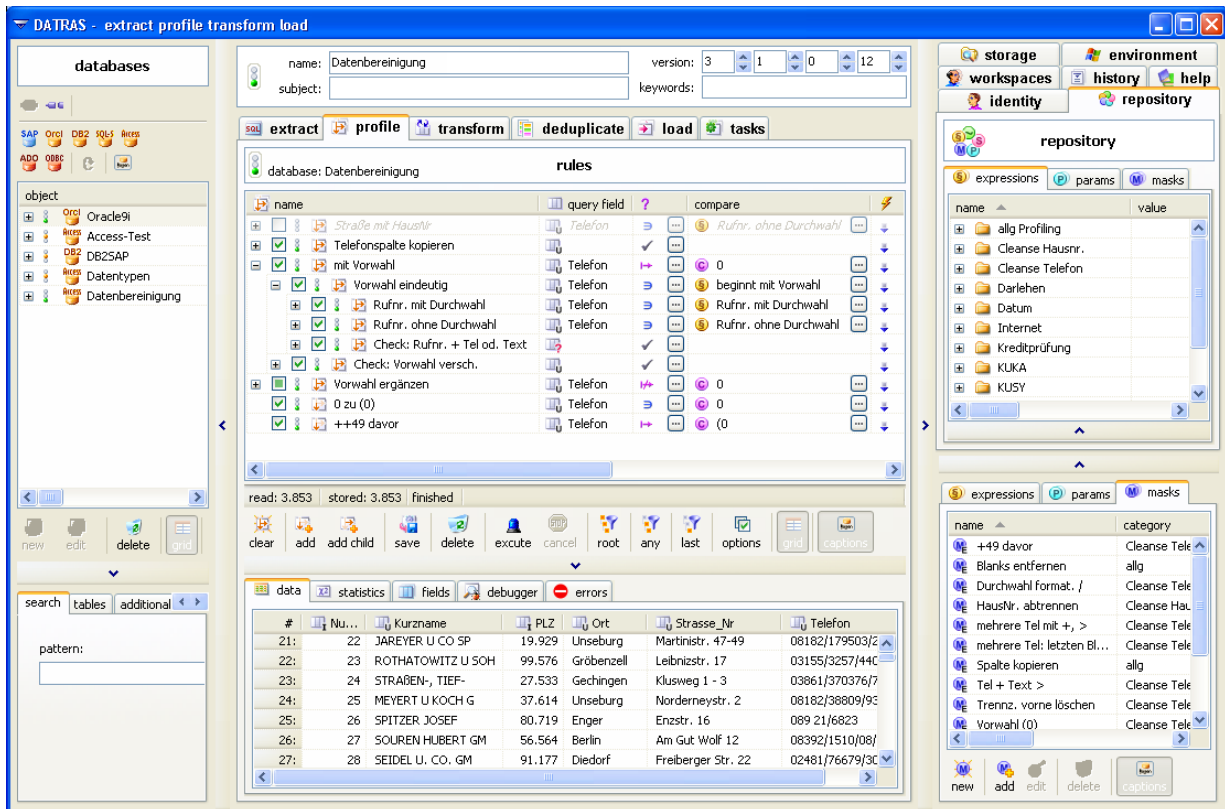


Datras EPTL

Extract – Profile – Transform – Load



Datras EPTL is an integrated solution which covers the complete process of data quality management (DQM), i.e. data analysis, data transformation and data integration of all sorts of company data. The basis for all activities are business rules that are

pre-built or user-defined, exchangeable and reusable. The integrated repository includes pre-built industry-specific rules and resources. Therefore user-specific DQM processes can be started directly with just few adjustments.

Features

- Native connectors for Oracle, IBM DB2, MS SQL-Server, MySQL, MS Access and SAP
- Text import with integrated data type conversion
- Explorative profiling incl. pattern recognition
- Transparent deduplication
- Debugger and editors
- Database explorer

Repository:

- Expressions, parameters, masks and compares
- Industry-specific business rules
- Completely pre-built data quality tasks as resources

Benefits

- High-performance database access
- Rules and resources are exchangeable and distributable as XML files (team collaboration)
- Easy and manageable realisation of user-defined customizing
- Visual representation of dependencies and meta information of company data
- Predefined components for data quality activities
- Easy definition of user-defined profiles and customizing of predefined profiles
- Minimal customizing needed for user-defined DQM processes

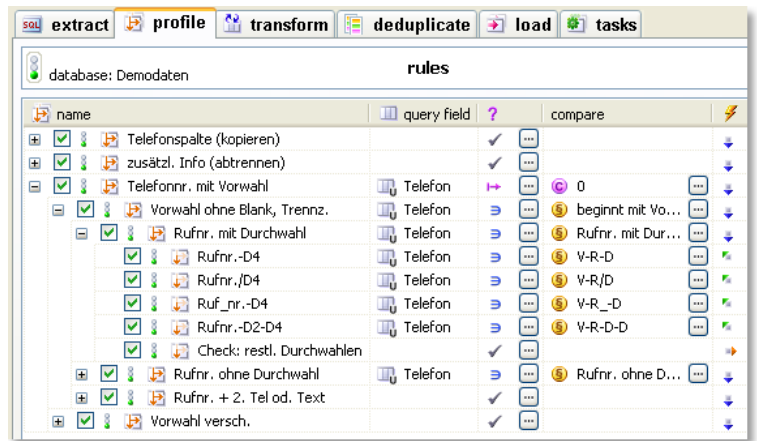
Extract & Profile

Extract

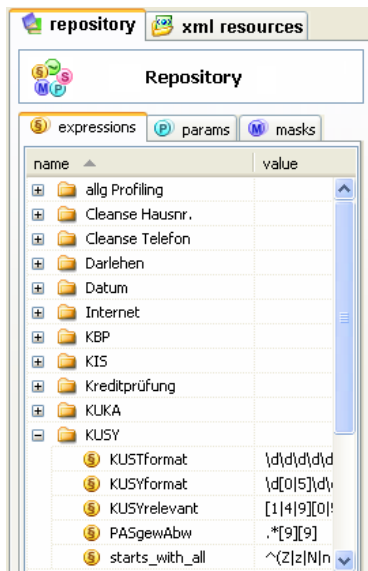
- Data extract for SQL databases as well as text and Excel data
- Wizards for creating SQL queries

Profile

- Real time profiling
- Standard profiling (min, max, average, etc.)
- Explorative profiling (pattern recognition, value and character distributions)
- Column, dependency, redundancy profiling based on hierarchical rules



Profiling result are linked to the source data and can thereby used as filters for subsequent data quality activities like transform, load etc.



Repository

- **Expressions:** Set of regular expressions as basic components for the business rules
- **Params:** Parameters for dynamic control of sequences, in particular the setting of time frames for Extract and Profile steps
- **Masks:** Masks for controlled transforming of data types and data contents
- **Rules:** Predefined and industry-specific business rules to create and customize user-defined profiles
- **Compares:** Predefined impreciseness profiles for detection of duplicates
- **Resources:** Complete sequences from Extract to Load with all required components

All components of the repository are exchangeable between users and can be unloaded and distributed as XML files.

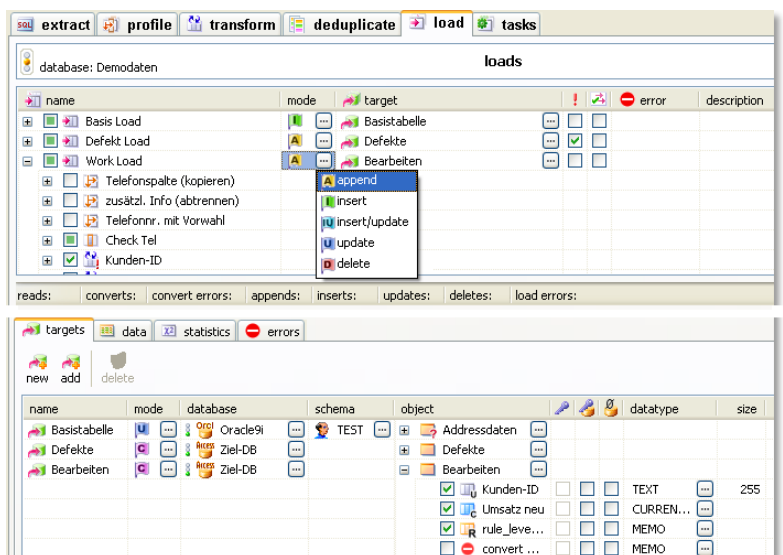
Transform & Load

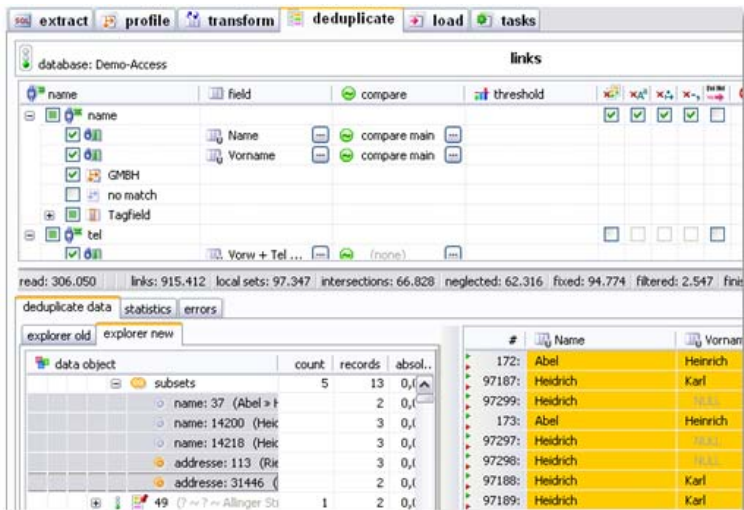
Transform

Based on the profiling results data types as well as data content can be manipulated (data transformation and data cleansing) using so called Converts. These Converts can be controlled using Masks which describe the desired transformation. In case of defective data supplementary information is added to the data records to help users identify and correct data anomalies.

Load

The transformed data can be filtered and distributed according to the profiling results and the transformation errors into different and independent SQL databases for subsequent processing.





Deduplication

Powerful and transparent doublet analysis which allows users to define the degree of impreciseness individually. In addition, the analysis can be extended to identify doublets based on information chains.

Compares

Compares allow the user to define the degree of impreciseness search individually and transparently and to save such impreciseness definitions in the repository for further analysis.

Tasks (optional)

Datras EPTL provides a client-server architecture to allow users to develop and maintain data rules decentralized and to operate complete data quality activities centralized on one or more servers using "Tasks".

This allows organisations to completely automate all their data quality activities on central server systems for production processing in an unattended operation manner.

