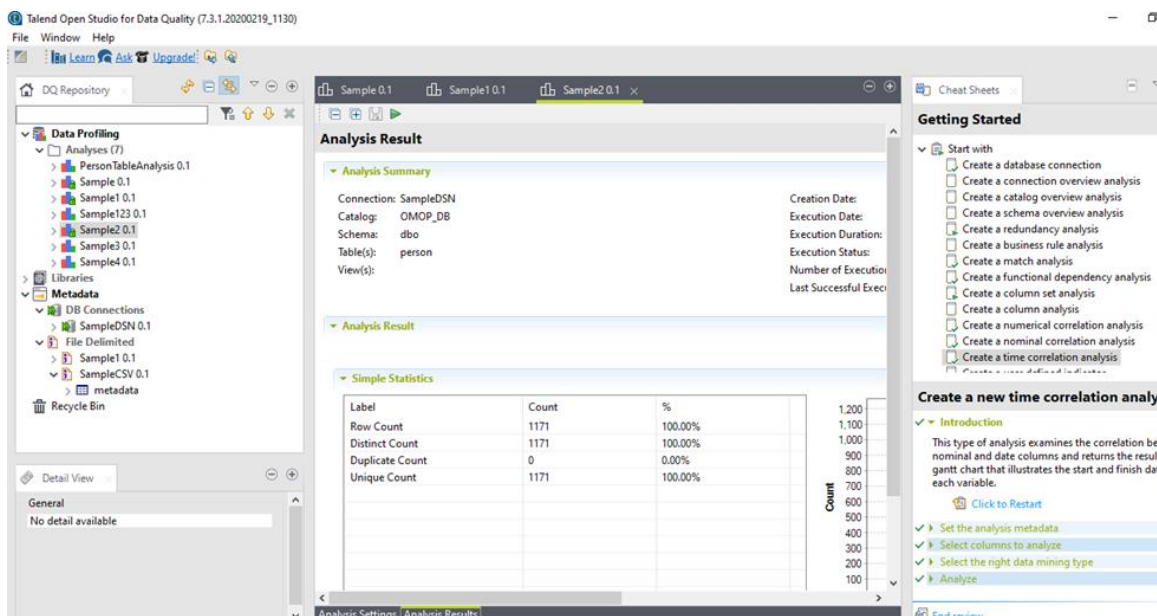




TALEND OPEN STUDIO FOR DATA QUALITY

(LIMITED FREE VERSION)

Assesses accuracy and integrity of data - Data Profiling Tool



Links

- ✓ <https://www.talend.com/products/talend-open-studio/>
- ✓ <https://www.talend.com/products/specifications-data-preparation/>
- ✓ <https://www.talend.com/services/training/catalog/data-preparation/>
- ✓ <https://www.talend.com/products/data-preparation-free-desktop-manuals-release-notes/>
- ✓ <https://www.talend.com/products/data-preparation/data-preparation-free-desktop/>
- ✓ https://download-mirror2.talend.com/tdp/user-guide-download/V251/TalendDataPrep_GettingStarted_EN.pdf

License Apache License 2.0
Version v2.5
Last Update 6/5/2018
OS macOS and Windows

System Requirements Processor 64-bit processor is required, Allocated memory 1GB minimum, Disk space 500MB minimum + datasets = 5GB recommended

Description Talend Open Studio for Data Quality is a freely downloadable and limited functionality version of the fuller featured product. Users will very quickly hit the tool's limits and require better connectivity to more data sources, and better matching, cleansing and visualization capabilities. Datasets can be a local or remote file that can be imported into the Talend Data Preparation tool (or from a database connection or other data sources, **although not in the context of the Free Desktop version**).

Features

- Import data

- Data standardization
- Basic formatting and cleansing
- Ability to add data cleaning recipes for reuse.
- Talend Data Preparation automatically suggests the proper semantic type for each column of your datasets.
- Data quality bar - under each column is a color-coded data quality bar that displays the number of fields that have correct data, empty fields, or incorrect data
- Basic text manipulation - clean and change values in a field with invalid values
- Basic numeric manipulation - clean and change outliers in a numeric field.
- Data masking - mask sensitive data
- Data blending - connect data from different sources. Take data from a preloaded dataset and add to a dataset you are working on
- Group and standardize - find cells with similar content and group them together by changes the text to match

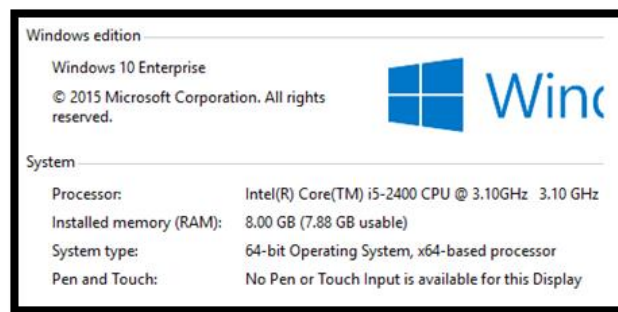
Connectivity / Supported Data Sources & Formats

- Local or remote file that can be imported into the Talend Data Preparation tool (or from a database connection or other data sources, although not in the context of the Free Desktop version).
- Excel or CSV file
- 90+ data sources and scale with Stitch Data Loader - <https://www.talend.com/products/pricing-model/>

Limitations

For an initial data investigation, which may include examining the existence and validity of attribute values, data ratios, uniqueness of keys, or orphaned records, the tool is "good enough." If you need to do more than this, you will need to full featured product.

Performance



The tool was able to load the 1.3M records and the memory consumption was high

TALEND OPEN STUDIO FOR DATA QUALITY - DATA ANALYSIS

1. Download TQS from "<https://www.talend.com/products/talend-open-studio/>"
2. Open "TOS_DQ-win-x86_64.Exe" to start the application.

Database Connection

Update Database Connection - Step 2/2

DB port must be specified

DB Type: Microsoft SQL Server

Db Version: Microsoft

String of Connection: jdbc:sqlserver://DESKTOP-SA188C6\SQLEXPRESS;DatabaseName=OMOP_DB;

Login: sa

Password:

Server: DESKTOP-SA188C6\SQLEXPRESS

Port:

DataBase: OMOP_DB

Schema: dbo

Additional parameters:

File - Step 2 of 3

Edit an existing Metadata File on repository
Update the path of the file and the format settings

File Settings

Server: Localhost 127.0.0.1

File: D:\OfficeFiles\2019_Caradigm\HDR-UK\synthea_sample_data_csv_apr2020\csv\patients.csv Browse...

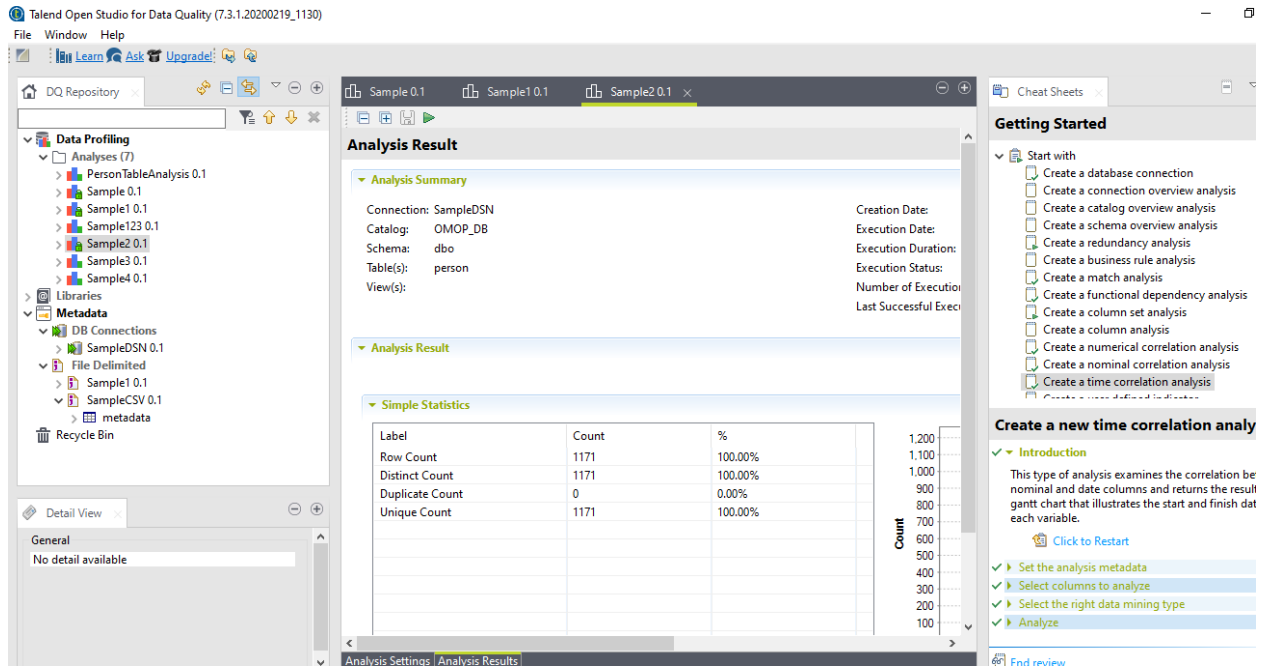
Format: WINDOWS

File Viewer

```
Id,BIRTHDATE,DEATHDATE,SSN,DRIVERS,PASSPORT,PREFIX,FIRST,LAST,SUFFIX,MAIDEN,MARITAL,RACE,ETHNICITY,GENDER,BIRTHPLACE,ADCE
1d604da9-9a81-4ba9-80c2-de3375d59b40,1989-05-25,,999-76-6866,S99984236,X19277260X,Mr.,José Eduardo181,Gómez206,,M,white,hispani
034e9e3b-2def-4559-bb2a-7850888ae060,1983-11-14,,999-73-5361,S99962402,X88275464X,Mr.,Milo271,Feil794,,M,white,nonhispanic,M,Danv
10339b10-3cd1-4ac3-ac13-ec26728cb592,1992-06-02,,999-27-3385,S99972682,X73754411X,Mr.,Jayson808,Fadel536,,M,white,nonhispanic,M,
8d4c4326-e9de-4f45-9a4c-f8c36bff89ae,1978-05-27,,999-85-4926,S99974448,X40915583X,Mrs.,Mariana775,Rutherford999,Williamson769,M,
f5dcd418-09fe-4a2f-baa0-3da800bd8c3a,1996-10-18,,999-60-7372,S99915787,X86772962X,Mr.,Gregorio366,Auer97,,white,nonhispanic,M,Pat
72c0b9ce-7aa4-430b-aaff-bd0ce7846e55,2017-07-27,,999-68-6630,,Jacinto644,Kris249,,white,nonhispanic,M,Beverly Massachusetts US,88
b1e9b0b9-da6e-4f68-b603-bd896a50ca86,2003-12-13,,999-73-2461,S99954048,,Jimmie93,Harris789,,white,nonhispanic,F,Worcester Massac
01207ecd-9dff-4754-8887-4652eda231e2,2019-05-15,,999-81-4349,,Karyn217,Mueller846,,white,nonhispanic,F,New Bedford Massachusetts
b58731cc-2d8b-4c2d-b327-4cab771af3ef,1970-05-16,,999-90-2484,S99978036,X78170348X,Mrs.,Isabel214,Lucio648,Carvajal694,M,white,hispa
cfef79fc-df05-476e-b274-43e09ea345db,2016-07-04,,999-15-5895,,Alva958,Krajcik437,,white,nonhispanic,F,Boston Massachusetts US,104
ad2e9916-4979-40fc-a8c0-68651a0cb5a6,2004-12-19,,999-78-4480,,Jeffrey461,Greenfelder433,,white,nonhispanic,M,Plymouth Massachus
bf6f6537h-535a-4f31-9a56-073220f96a17,1991-07-03,,999-74-9712,S99913545,X65838399X,Mrs.,Karyn217,Isact432,Revnndlc644,M,white,nonhi
<
```

< Back Next > Finish Cancel

SAMPLE TQS STATISTICS:



LIMITATIONS:

- The tool was able to load the 1.3M records and the memory consumption was high.