Data Cleanser

for Quarterly Skills Development Reports

USER GUIDE

Edition 1 (for Data Cleanser, version 1.00)

CONTENTS

Abbreviations and Acronyms

App.   - Computer Application
GUI   - Graphical User Interface
QSDR - Quarterly Skills Development Report
SA ID - South African Identity Document

1. Introduction

1.1. Purpose of application

The purpose of this application is to automate the process of removing errors from *Quarterly Skills Development Report*s (QSDR)

A QSDR is an excel document containing information on all training programmes offered to employees of an organisation for a particular financial quarter. This document is used to plan budgets for skills training.

Due to its financial implications, its imperative that any errors are removed from these documents. Manually verifying information in QSDRs is a lengthy, tedious and error-prone process.

This application automatically detects errors, and corrects/removes them.

1.2. How do I use this app?

This section briefly describes how to use *Data Cleanser* as well as some detail on how the app works.

Data Cleanser is 'data scrubber', which is an application that corrects incorrect or incomplete data. In this instance, it remedies data residing within QSDR reports.

The user selects QSDR excel file(s) to be cleansed by clicking on *Find file*. Thereafter, cleansing is initiated by clicking on *Cleanse Data*.

The *Data Cleanser* iterates through each excel file within the application's *Input files* directory, and iterates through each sheet within each excel workbook to employ cleansing mechanisms. The application uses the South African ID number as an identifier. Data from Persal and Training Intervention 'databases' are used to detect and correct erroneous QSDR report data.

After the input QSDR report(s) are inspected and corrections are applied where applicable, new 'cleansed' output QSDR report(s) are produced by *Data Cleanser*, and can be found in the *Output files* directory. This directory is automatically opened by the applicaiton once cleansing is complete. This directory can also be accessed by clicking on the *Output files* button on the application's Graphical User Interface (GUI).
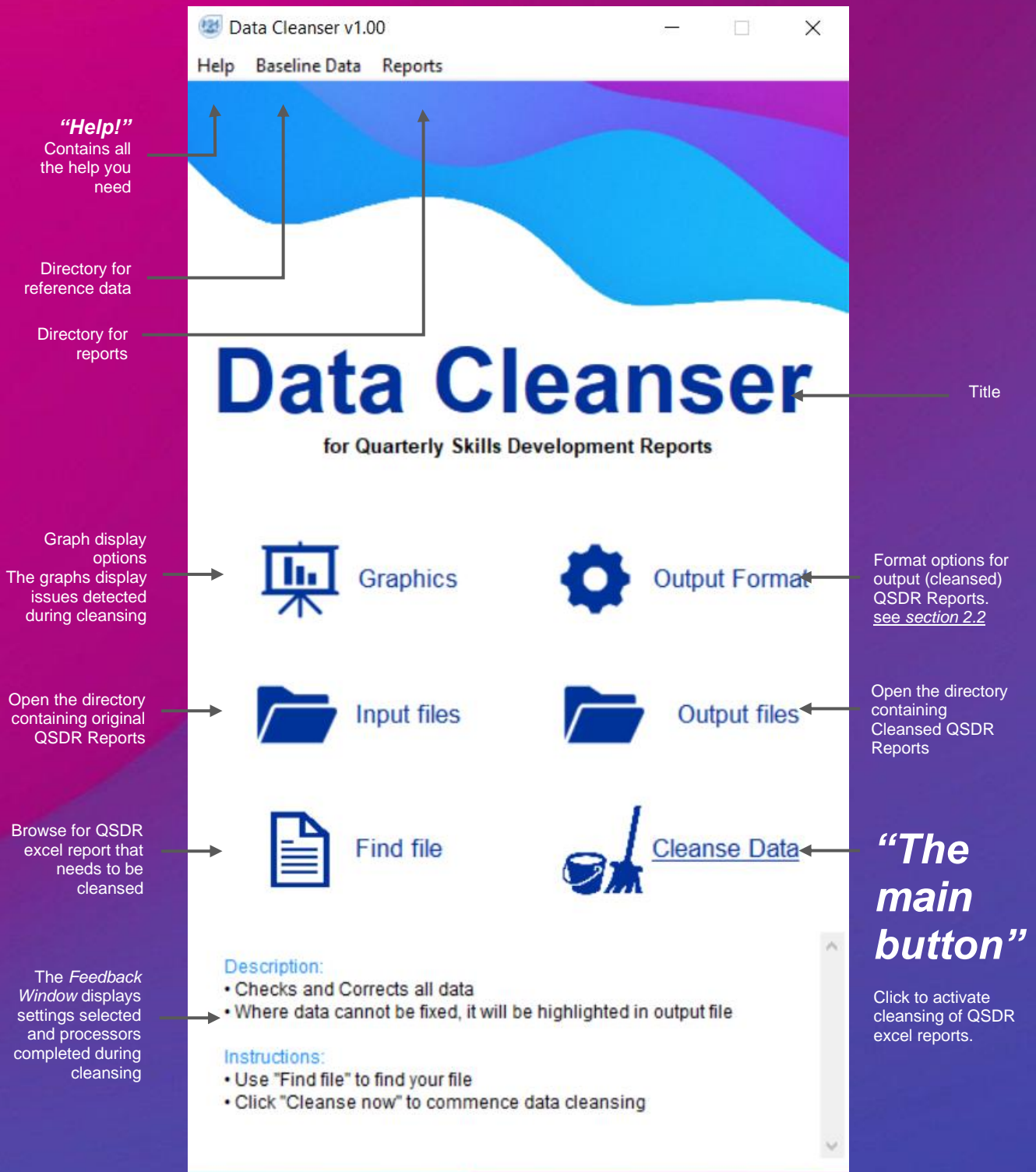
Additionally, reports on issues detected and/or fixed by *Data Cleanser* are produced after each instance of use. These reports assume the form of a bar graph diagram and an excel report which can be located in the *Result Diagrams* and *Result Report* respective directories.

The Result Reports inform the user on any residual issues that exist within the cleansed output QSDR reports. Residual issues appear as colour-coded ID numbers in the output reports. This is explained in greater detail under section 3.3.

2. What is this button?

This Chapter provides more detail on how to use *Data Cleanser*.

2.1. Interface

"Help!"
Contains all the help you need

Directory for reference data

Directory for reports

Title

Graph display options
The graphs display issues detected during cleansing

Format options for output (cleansed) QSDR Reports.
see *section 2.2*

Open the directory containing original QSDR Reports

Open the directory containing Cleansed QSDR Reports

Browse for QSDR excel report that needs to be cleansed

"The main button"

Click to activate cleansing of QSDR excel reports.

The *Feedback Window* displays settings selected and processors completed during cleansing

---

Data Cleanser v1.00

Help   Baseline Data   Reports

Data Cleanser
for Quarterly Skills Development Reports

Graphics   Output Format

Input files   Output files

Find file   Cleanse Data

Description:
• Checks and Corrects all data
• Where data cannot be fixed, it will be highlighted in output file

Instructions:
• Use "Find file" to find your file
• Click "Cleanse now" to commence data cleansing

2.2. Output Format

Provides excel format options for cleansed output QSDR file

OPTION	EXPLANATION
Formatted Workbook	Cleansed output QSDR report looks identical to the input report, except that it's clean! This is the ideal option. Creates an .xlsm file which runs a VBA macro that in turn combines all slave worksheets into a master workbook. Thereafter, it converts the .xlsm file into a .xlsx file. **Warning**: This Option only works if "all macros" are enabled within Excel! To enable this function, do the following in Excel: *File* -> *Options* -> *Trust Center* -> *Trust Center Settings* -> *Macro Settings* -> *Enable all macros*
Simple Workbook	Provides cleansed .xlsx files with none of the formatting that exists in the original QSDR workbooks This format would be adequate for adoption of cleansed QSDR data into the *Workplace Skills Plan and Quarterly Monitoring Systems* website (once it is complete)
Separate Sheets	This is the default option. **Note**: Each worksheet of the original QSDR file is outputted as an individual .xlsx file! Each outputted file would have the correct font formatting.

2.3. Graphics

This menu provides options for graphical representations of errors found and corrected during cleansing


| OPTION	EXPLANATION		
Output Figure	Choose between bar graph or pie chart display of errors in original QSDR reports		
Output State	Activate	Automatically displays graph of errors once data cleansing complete	
	Deactivate	Deactivates the functionality described above	
Open Location	Opens the directory containing the error graphs that are automatically saved after every instance of use.		

2.4. Dropdown menus

Describes the function of each dropdown menu

Help
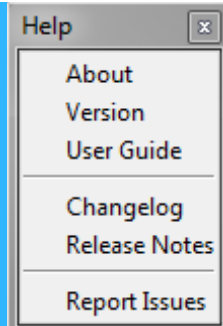
|	
About	Brief description of the application
Version	States the version of the application
User Guide	Opens this pdf document
Release Notes	Latest changes applied to application
Report Issues	Instructions on how users can report issues to support team

Settings
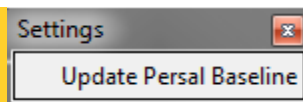
Update reference/ baseline Persal data

N.B! It is imperative that the Persal data is up-to-date to ensure sound QSDR data cleansing. Moreover, do not upload an incorrect file!

Other
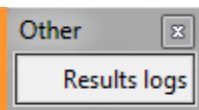
Opens directory containing 'Results logs'. These are excel documents containing information on the errors found in the original QSDR excel workbooks.
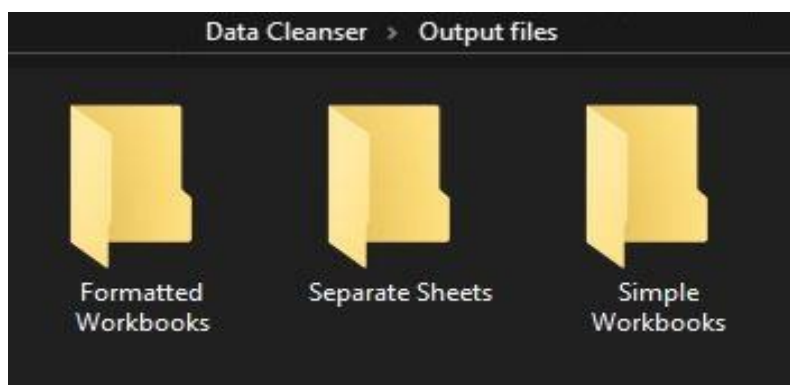
3. Output

This chapter provides the reader with an understanding of the application's output.

3.1. Cleansed Excel Workbooks directory

All cleansed excel workbooks will be found in the *Output files* folder within the application's directory. In turn, the folder within *Output files* that the cleansed workbooks resides depends on the *Output Format* selected before data cleansing

Folder Name (in *Output Files*)	*Output Format* (selected)
Formatted Workbooks	Formatted Workbook
Simple Workbooks	Simple Workbook
Separate Sheets	Separate Sheets

Figure: Location of cleansed excel documents

3.2. Data cleansing reports

In addition to the cleansed Excel workbooks, the application outputs reports which provide the user with information, on the data cleansed as well as the data that requires additional intervention, by means of Excel documents that can be found in the *Result Reports* folder. This folder can be accessed within the application by clicking on *Result logs* under the *Other* dropdown menu on the application's interface.

Moreover, graphs describing some of the issues resolved during cleansing are produced by the application and can be found in the *Result Diagrams* folder. This too can be accessed from within the application (*Graphics* button -> *Open Location of Graphics*).

3.3. Colour coded residual issues

If *VBA* or *Separate* option was selected as the *Output Format* then any residual issues (data that could not be verified and/or corrected by the application) will be colour coded in the **output** workbook(s). In turn, the colour will depend on the nature of the issue:

Colour	Description
Red	**Gross errors** that cannot be verified
Orange	Data does not exist in *Persal* database. Therefore cannot be verified.
Yellow	Data does not exist in *Training Interventions* database. Therefore cannot be verified.
Pink	Where value for **disability** differs to corresponding values in Persal database.

4. Baseline Data

To view all databases, go to **DB** (database) directory
Baseline Data (Dropdown menu) > Open Baseline Data Directory

4.1. Update Baseline data

1) Go to **DB** directory

2) Remove/ Overwrite existing database(s)

3) Ensure that the name of the new database file is correct:
   • **Persal_DB** for Persal Database
   • **Training_DB** for Training Interventions Database

4) Ensure that each database file has the correct sheet name:
   • **Persal** for Persal Database
   • **Training** for Training Database

5) Ensure that each database file has the correct column names:

column names for **Persal_DB**	column names for **Training_DB**
ID Number Today Persalno Surname FIRST NAMES Race SL-Person Job Title on Rank Disabled Occ Cat Labour ADDRESS 3	Intervention NQF Level Type Duration Unit Of Measure Provider / Source

Note: All names (for files, sheets, and columns) are case sensitive. Ensure that all names are shown exactly as they appear in above.