

The European Commission's science and knowledge service

Joint Research Centre



JRC-ENCR Data Quality Check Software: introduction

Francesco Giusti and Carmen Martos
European Commission, Joint Research Centre (JRC),
Ispra (VA), Italy

ENCR-JRC Training on Cancer Registry Data Coding

Ispra, 5 June 2019

A proposal on cancer quality checks: one common procedure for European cancer registries

ENCR-JRC Working Group



RARECARE/
RARECARENet

ENCR

IARC

JRC

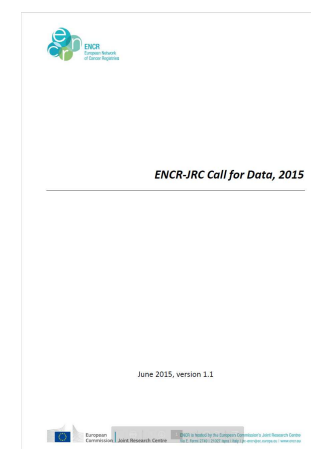
Experts from the
cancer registry
community

CONCORD

EUROCARE



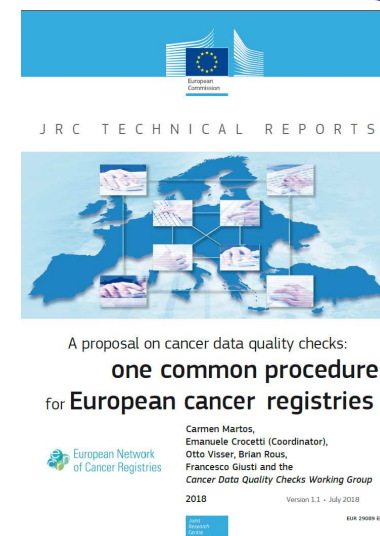
2014



2015



JRC-ENCR QCS



2018

4 Quality Checklist for Multiple Primary Malignant Tumours

The quality checklist of warnings for Multiple Primary Malignant (MPMTs) was developed by the JRC according to the current International Rules for Multiple Primary Cancers published in 2004 (http://www.encyr.eu/images/docs/recommendations/MPRules_july2004.pdf). This checklist has been addressed by the JRC at a later stage than for version 1.0 of this report, and is therefore included in the current version 1.1 to clarify the checks and the warning messages given by the JRC-ENCR Cancer Registries Data Quality Check Software. The steps for checking solid MPMTs

are the following:
Step 1. The two topographies are compared according to the current International Rules for Multiple Malignant Cancers published in 2004. In addition to the groups of topography codes considered as a single site in table 1 of the 2004 international rules, for checking other groups are considered as a single topography (see Table 9).
C80 (unknown primary site) and C768 (overlapping lesion of ill-defined sites) are considered as a single site with any topography.

Table 9. Groups of topography codes considered as a single site for solid tumours

Topography code	Definition
C00	Lip
C03	Gum
C04	Floor of mouth
C05	Palate
C06	Other and unspecified parts of mouth
C760	Head, face or neck NOS

The JRC-ENCR Quality Check Software



JRC TECHNICAL REPORTS

The JRC-ENCR Quality Check Software (QCS) for the validation of cancer registry data: user compendium

JRC-ENCR QCS Version 1.8
1.8.1 update

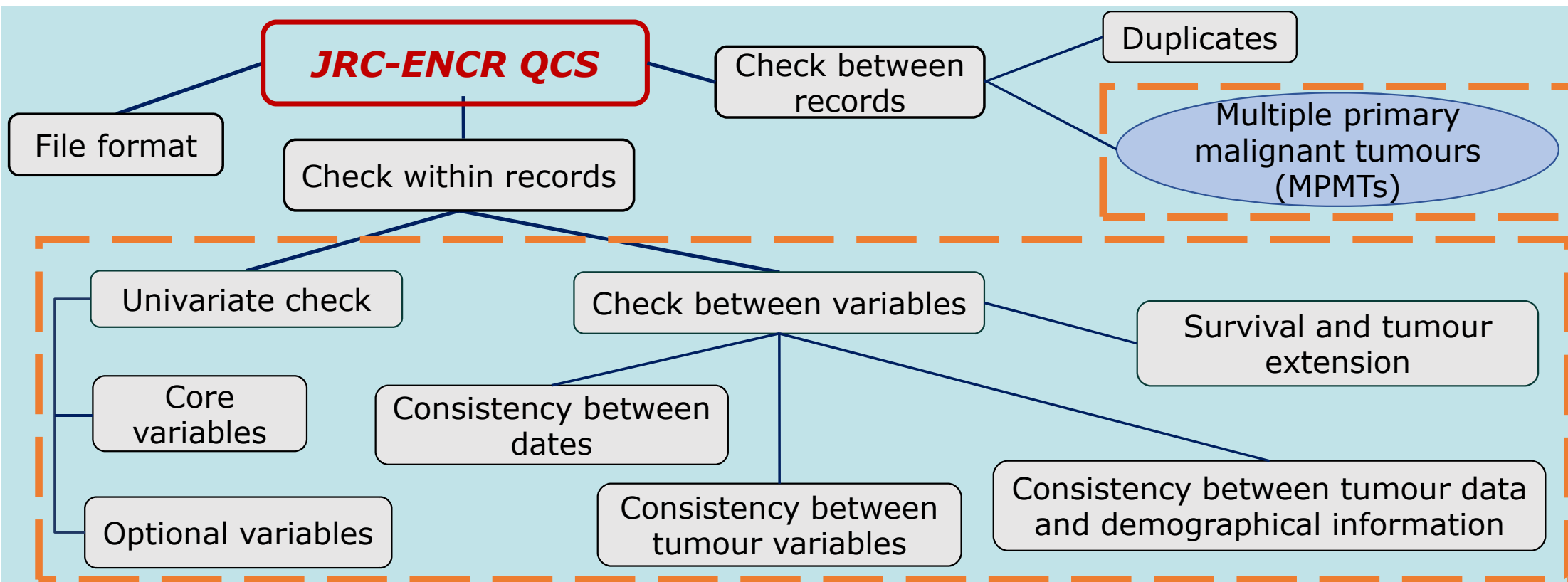
Francesco Giusti, Carmen Martos, Stefano Adriani, Tadeusz Dyba, Lena Voith von Voithenberg, Luciana Neamtii, Raquel N. Carvalho, Giorgia Randi, Nadya Dimitrova, Nicholas Nicholson, Revveka Trigka, Enrico Ben, Emanuele Crocetti, Manola Bettio

2019

- The JRC-ENCR Quality Check Software (QCS) was created for checking the **internal consistency** of cancer-registries' data against the requirements of the 2015 protocol of the *ENCR-JRC Call for data*
- **THE INSTALLATION OF JAVA IS NECESSARY FOR RUNNING THE QCS**



The JRC data quality check process



From version 1.7 to version 1.8

- Many cancer registries provided feedback on the QCS
- Improvements and better flexibility following comments from registries
- Request to implement the JRC-ENCR QCS in the NICERStat Swiss system
- Similar requests from other countries: Norway, Finland, Germany, Austria
- Our experience using the QCS for checking the data submitted by more than 150 European cancer registries

What's new in version 1.8

- **Multi-Platform:** possibility to use the QCS in **macOS** and **Linux** operating systems (see sections 2.5 and 2.6 below)

2.5 Running the QCS on macOS

1. Double click the ZIP file: the package will be unzipped in a new folder, having the same name of the ZIP package (but without any extension)
2. Press the combination *Command-Shift-U* (Command is the key with the Mac symbol) to open the Utility window
3. Double click the Terminal icon (or label, depending by your view settings) to open a Terminal window
4. Enter the Terminal window and move into the folder created at **step 1**. For example, if the target QCS file was named "JRC-ENCR-QCS-V1.8.zip", then you should execute the command:

```
cd Desktop/JRC-ENCR-QCS-V1.8
```

5. Execute the file having the extension ".sh". For example if the file is named "start-jrc-encr-qcs.sh", then type the command:

```
./start-jrc-encr-qcs.sh
```

2.6 Running the QCS on Linux operating systems

1. Unzip the ZIP file into the directory where to wish to install the application. For example, if the target QCS file was named "JRC-ENCR-QCS-V1.8.zip" you should execute the command

```
unzip JRC-ENCR-QCS-V1.8.zip
```

2. Move to the folder created at step 1. For example:

```
cd JRC-ENCR-QCS-V1.8/
```

3. Make sure the ".sh" file has permissions for execution. If not, assign it executable permissions by typing the command:

```
chmod +x start-jrc-encr-qcs.sh
```

4. Execute the QCS by running the ".sh" file:

```
./start-jrc-encr-qcs.sh
```


What's new in version 1.8

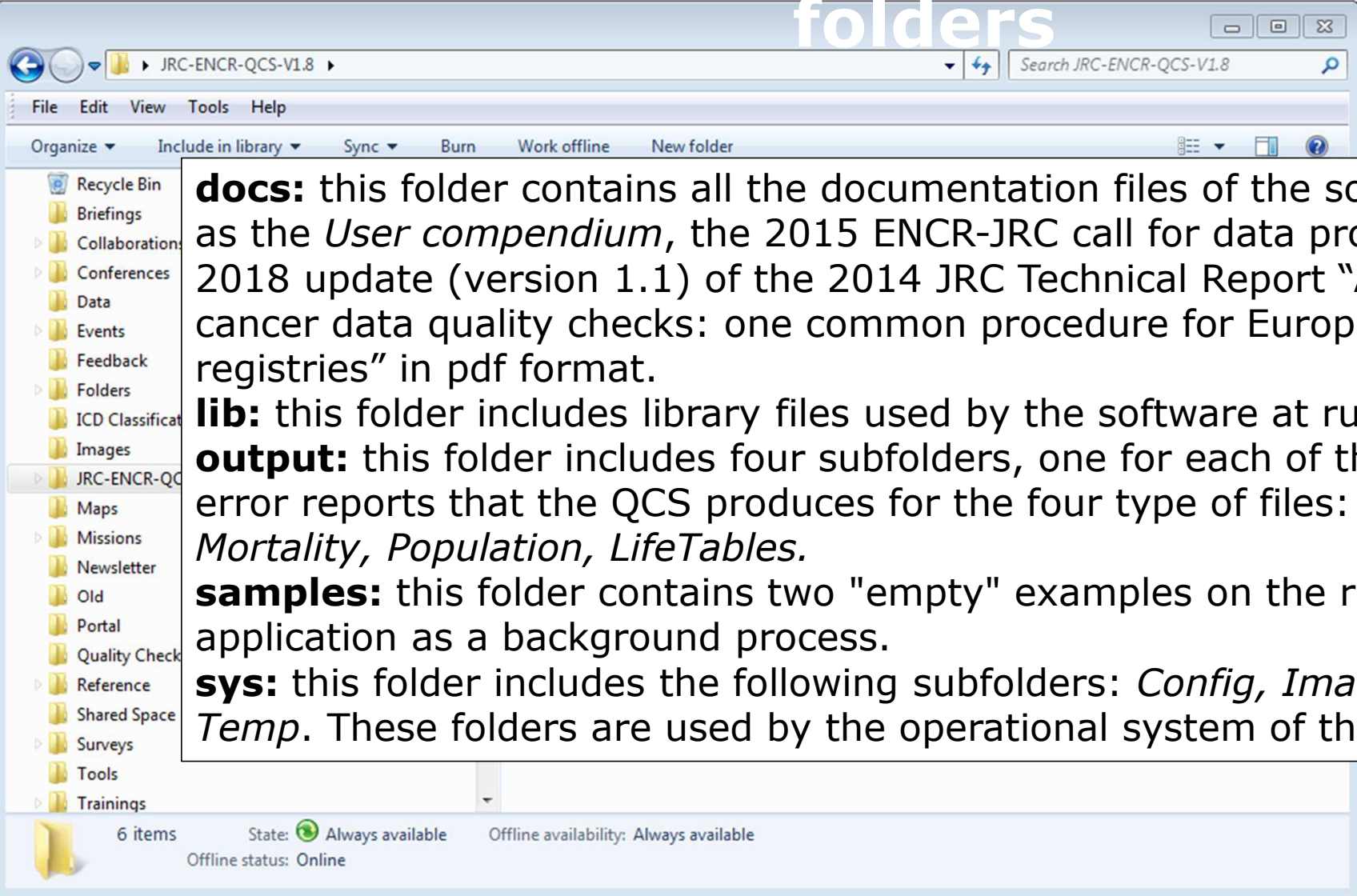
- Creation of a **separate file** (***QCS-PrimaryDuplicates-Output.csv***) for Multiple primary malignant tumours

	A	B	C	D	E	F	G	H	I	J	K	L
1	2_Patient_ID	3_Tumour_ID	12_BoD	13_Topo	14_Morph	15_Beh	7_Sex	DoI	DoB	18_Vital_Status	23_Laterality	Error_code
2	1	1	7	C809	8070	3	1	99/3/1983	99/10/1926	1		W-MPMT
3	1	2	7	C090	8071	3	1	99/8/1999	99/10/1926	1	2	W-MPMT
4	2	1	7	C447	8721	3	2	99/12/198	99/9/1920	1	9	W-MPMT
5	2	2	2	C809	8000	3	2	99/8/2008	99/9/1920	1		W-MPMT
6	3	1	7	C672	8120	3	1	99/7/2003	99/4/1941	1		W-MPMT
7	3	2	7	C809	8020	3	1	99/2/1990	99/4/1941	1		W-MPMT
8	4	1	7	C100	8070	3	1	99/4/2011	99/12/1949	1		W-MPMT
9	4	2	7	C760	8070	3	1	99/10/200	99/12/1949	1		W-MPMT
10	5	1	1	C809	8000	3	1	99/9/2000	99/7/1920	1		W-MPMT
11	5	2	7	C446	8720	3	1	99/4/1997	99/7/1920	1	9	W-MPMT
12	6	1	7	C129	8070	3	1	99/11/200	99/7/1943	1		W-MPMT
13	6	2	7	C809	8020	3	1	99/3/1998	99/7/1943	1		W-MPMT
14	7	1	7	C809	8010	3	1	99/1/1986	99/1/1914	1		W-MPMT
15	7	2	7	C619	8140	3	1	99/3/1985	99/1/1914	1		W-MPMT
16	8	1	7	C760	8070	3	1	99/9/1999	99/2/1946	1		W-MPMT
17	8	2	7	C672	8120	3	1	99/10/201	99/2/1946	1		W-MPMT
18	9	1	7	C619	8140	3	1	99/5/1999	99/2/1928	1		W-MPMT
19	9	2	7	C809	8020	3	1	99/9/2006	99/2/1928	1		W-MPMT

What's new in version 1.8

- Update of the morphology groupings used for the MPMTs according to the **2011 ICD-O-3.1** classification
- Improvement of the MPMTs checks algorithms
- Possibility to mark whether and which configuration files have been modified by the user
- Improvement in the QCS flexibility, for instance on the formats of variables **30_pT, 32_pM, 33_cT, 35_cM**

JRC-ENCR Quality Check Software: folders



docs: this folder contains all the documentation files of the software, such as the *User compendium*, the 2015 ENCR-JRC call for data protocol and the 2018 update (version 1.1) of the 2014 JRC Technical Report "A proposal on cancer data quality checks: one common procedure for European cancer registries" in pdf format.

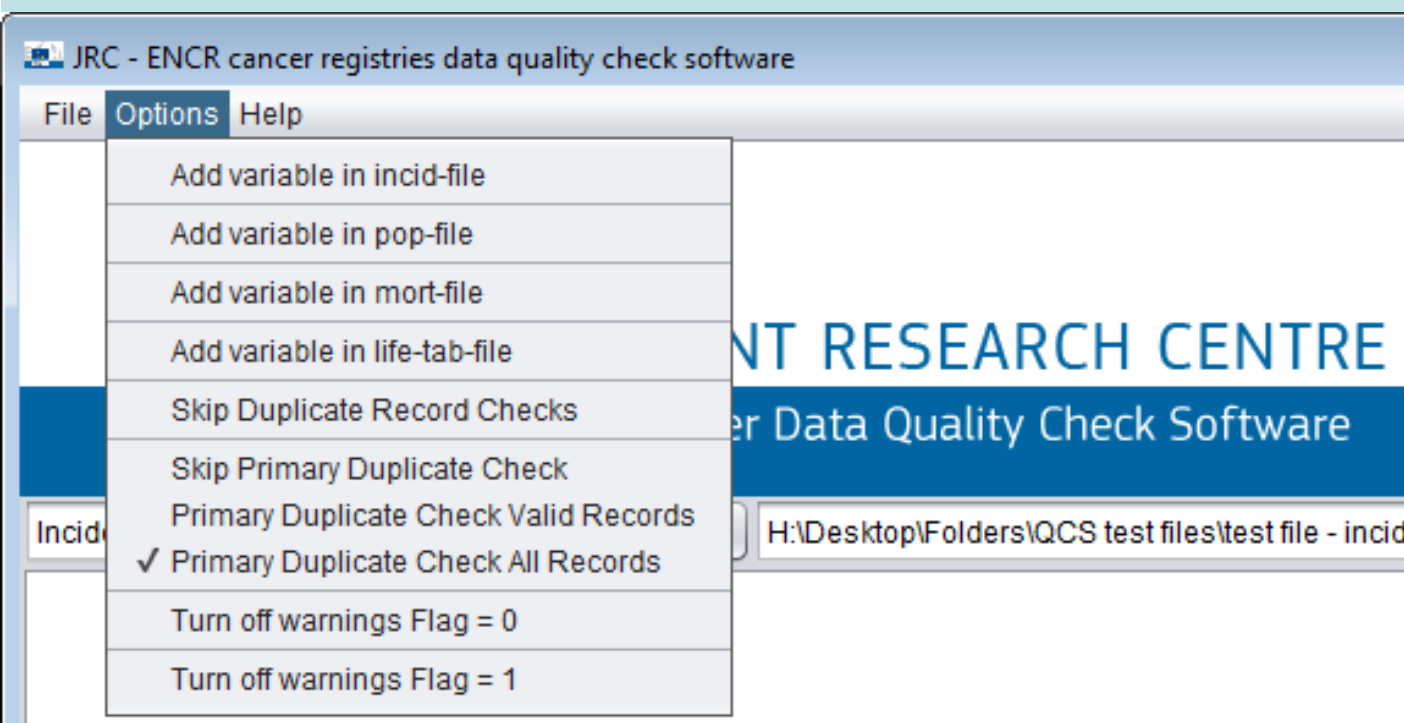
lib: this folder includes library files used by the software at run-time.

output: this folder includes four subfolders, one for each of the different error reports that the QCS produces for the four type of files: *Incidence, Mortality, Population, LifeTables*.

samples: this folder contains two "empty" examples on the running of the application as a background process.

sys: this folder includes the following subfolders: *Config, Images, Log, Temp*. These folders are used by the operational system of the QCS.

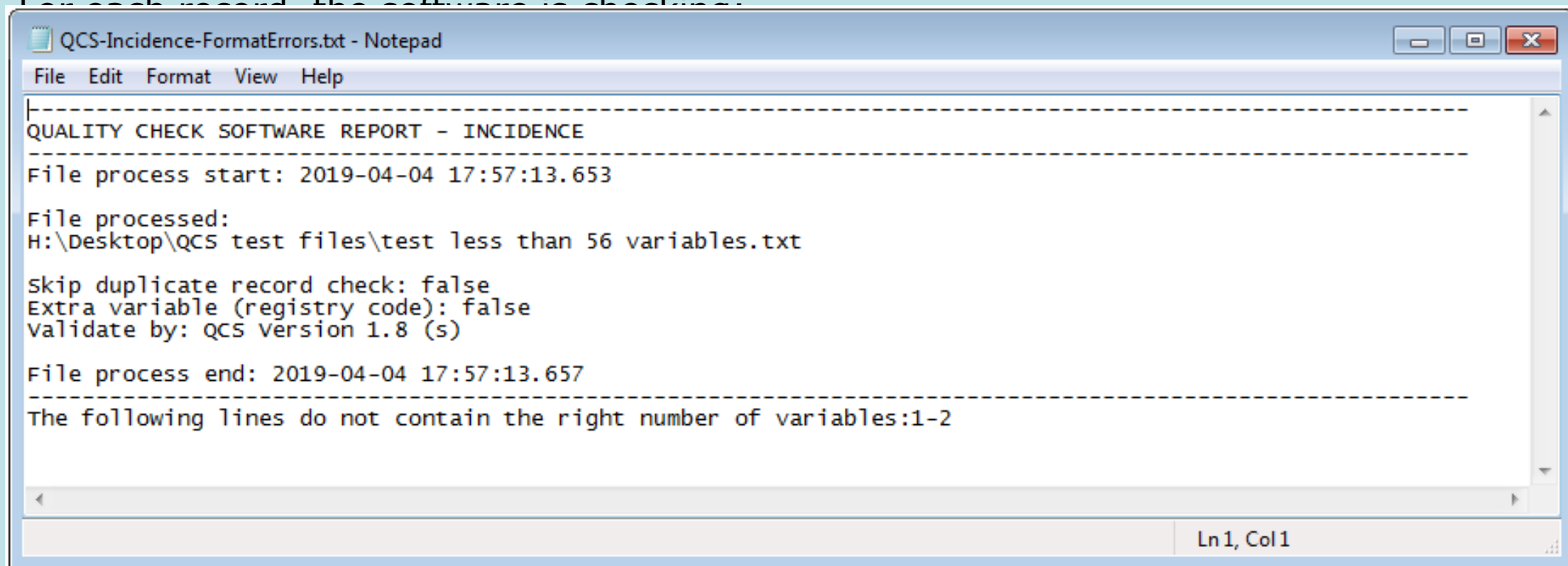
Running the JRC-ENCR Quality Check Software



- Prepare the file (header, number of variables, file format)
- Run (possibility to select different options)

JRC-ENCR Quality Check Software: file format check

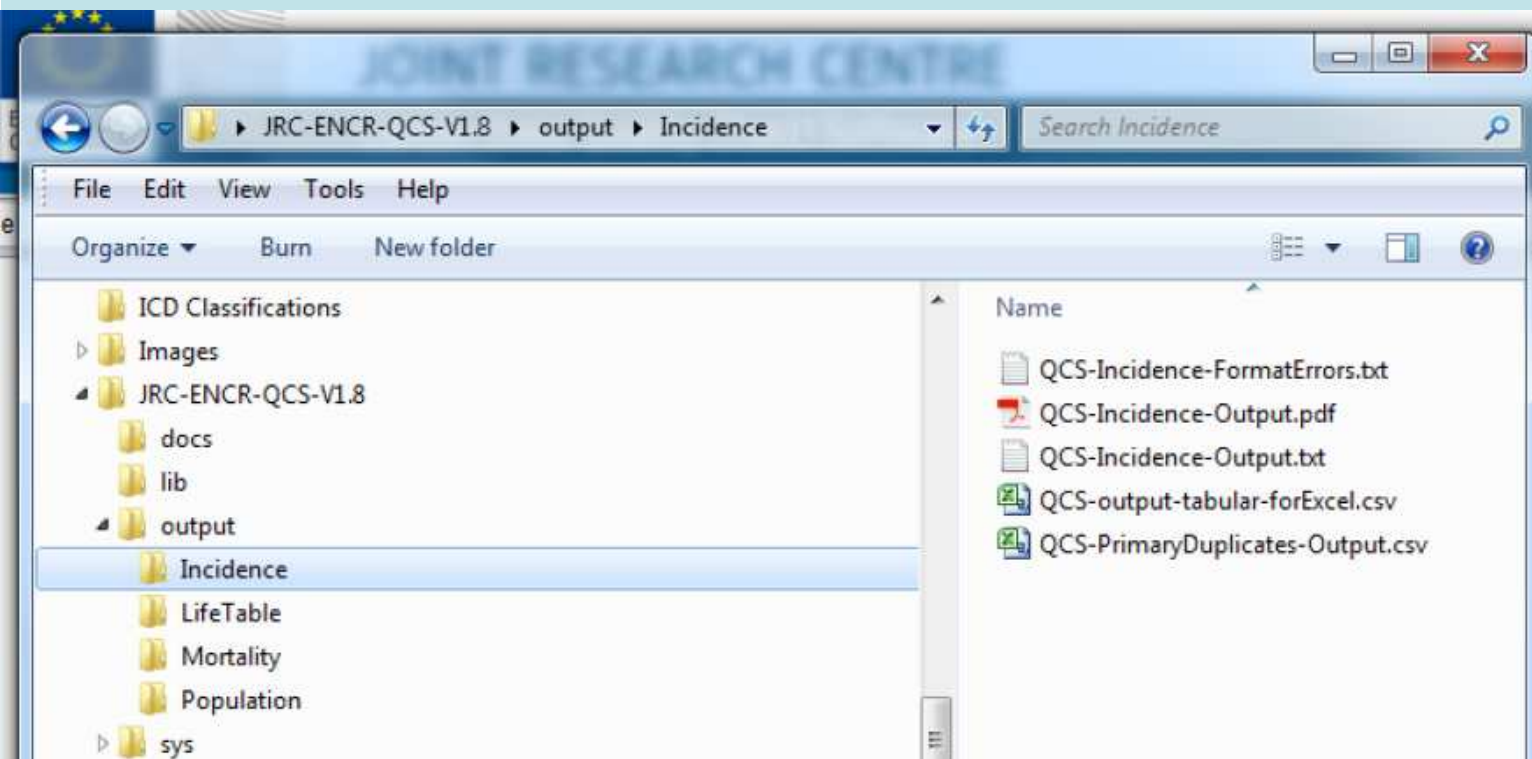
For each record, the software is checking:



The screenshot shows a Notepad window titled "QCS-Incidence-FormatErrors.txt - Notepad". The menu bar includes "File", "Edit", "Format", "View", and "Help". The text content is as follows:

```
-----  
QUALITY CHECK SOFTWARE REPORT - INCIDENCE  
-----  
File process start: 2019-04-04 17:57:13.653  
  
File processed:  
H:\Desktop\QCS test files\test less than 56 variables.txt  
  
skip duplicate record check: false  
Extra variable (registry code): false  
validate by: QCS Version 1.8 (s)  
  
File process end: 2019-04-04 17:57:13.657  
-----  
The following lines do not contain the right number of variables:1-2  
  
-----  
Ln 1, Col 1
```

JRC-ENCR QCS output: files



- **Format Errors**
- **PDF Output**
- **TXT Output**
- **CSV Output**
- **MPMTs Output**

JRC-ENCR QCS output: PDF and TXT files

PDF and **TXT** outputs (**QCS-Incidence-Output.pdf** and **.txt**)

Variables: 2_Patient_ID, 3_Tumour_ID, 1_Flag, 13_Flag, 14_Morpho, 15_Beh, 7_Sex, Doi, Dob, Var_Name, Var_Value, Error_Code

```
*****
ERRORS AND WARNINGS
*****
```

```
-----
2_Patient_ID                               3_Tumour_ID
-----
1_Flag  13_Topo  14_Morpho  15_Beh  7_Sex  DoI        DoB        Var_Name        Var_Value  Error_Code
-----
1        C421    9731      3      2      4/11/2014  2/12/1958  14_Morpho      9731      W-MOTO
                                           13_Topo      C421      W-MOTO
```


JRC-ENCR QCS output: CSV file

CSV output (QCS-output-tabular-forExcel.csv)

Variables: Line_nr, 2_Patient_ID, 3_Tumour_ID, 1_Flag, 13_Topo, 14_Morpho, 15_Beh, 7_Sex, Doi, Dob, Error_code, Error_Description, Var1_Name, Var1_Value, Var2_Name...

Line_nr	2_Patient_ID	3_Tumour_ID	1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	Doi	Dob	Error_code	Error_Description	Var1_Name	Var1_Value	Var2_Name	Var2_Va
209	13198	1	1	C421	9731	3	2	04/11/2014	02/12/1958	W-MOTO	Morphology + Topography not valid	13_Topo	C421	14_Morpho	9731
213	13490	1	1	C539	8000	3	2	30/06/2014	26/06/1970	W-BDMS	Morphology not specific enough (p.30)	14_Morpho	8000	12_BoD	7
217	13498	1	1	C445	8090	3	1	29/03/2014	31/05/1967	W-TOLA	Topography + Laterality not valid	13_Topo	C445	23_Laterality	3
251	13555	2	1	C445	8092	3	1	17/08/2014	10/10/1972	W-TOLA	Topography + Laterality not valid	13_Topo	C445	23_Laterality	3
444	13787	1	1	C445	8743	2	1	09/10/2014	21/10/1953	E-MOBE	Morphology + Behavior not valid	14_Morpho	8743	15_Beh	2
874	14002	1	1	C445	8743	2	1	10/11/2014	10/10/1952	W-TOLA	Topography + Laterality not valid	13_Topo	C445	23_Laterality	2
1903	15011	1	1	C421	9761	3	1	15/09/2015	23/11/1969	W-MOTO	Morphology + Topography not valid	13_Topo	C421	14_Morpho	9761
1951	15077	1	1	C445	8743	2	2	19/09/2015	02/03/1947	E-MOBE	Morphology + Behavior not valid	14_Morpho	8743	15_Beh	2
2566	15701	1	1	C421	9960	3	2	01/11/2015	14/03/1948	W-BDMS	Morphology not specific enough (p.30)	14_Morpho	9960	12_BoD	5
2571	15709	1	1	C445	8090	3	2	10/10/2015	27/03/1943	W-TOLA	Topography + Laterality not valid	13_Topo	C445	23_Laterality	2
2575	15722	1	1	C421	9962	3	1	23/09/2015	18/01/1934	W-BDMU	BoD + Morpho/Beh (p.30)	14_Morpho	9962	12_BoD	6
2756	15929	1	1	C421	9731	3	1	12/08/2015	15/08/1933	W-MOTO	Morphology + Topography not valid	13_Topo	C421	14_Morpho	9731

JRC-ENCR QCS output: MPMTs file

CSV output (QCS-PrimaryDuplicates-Output.csv)

Variables: 2_Patient_ID, 3_Tumour_ID, 12_Bod, 13_Topo, 14_Morpho, 15_Beh, 7_Sex, Doi, Dob, 18_Vital_status, 23_Laterality, Error_code

	A	B	C	D	E	F	G	H	I	J	K	L
1	2_Patient_ID	3_Tumour_ID	12_BoD	13_Topo	14_Morph	15_Beh	7_Sex	Doi	DoB	18_Vital_Status	23_Laterality	Error_code
2	1	1	7 C809	8070	3	1	99/3/1983	99/10/1926	1			W-MPMT
3	1	2	7 C090	8071	3	1	99/8/1999	99/10/1926	1		2	W-MPMT
4	2	1	7 C447	8721	3	2	99/12/198	99/9/1920	1		9	W-MPMT
5	2	2	2 C809	8000	3	2	99/8/2008	99/9/1920	1			W-MPMT
6	3	1	7 C672	8120	3	1	99/7/2003	99/4/1941	1			W-MPMT
7	3	2	7 C809	8020	3	1	99/2/1990	99/4/1941	1			W-MPMT
8	4	1	7 C100	8070	3	1	99/4/2011	99/12/1949	1			W-MPMT
9	4	2	7 C760	8070	3	1	99/10/200	99/12/1949	1			W-MPMT

Thank you



**ENCR-JRC Training on Cancer
Registry Data Coding**

JRC Ispra, 5-6 June 2019

JRC-ENCR@ec.europa.eu
<https://www.encr.eu/>