European Commission

The JRC – ENCR Cancer Registries Data Quality Check Software (QCS)

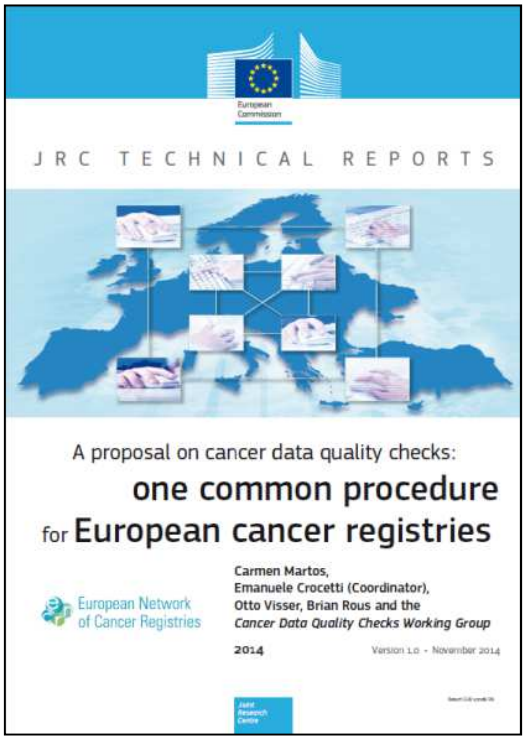
Francesco Giusti, JRC

Training on quality of cancer registry data

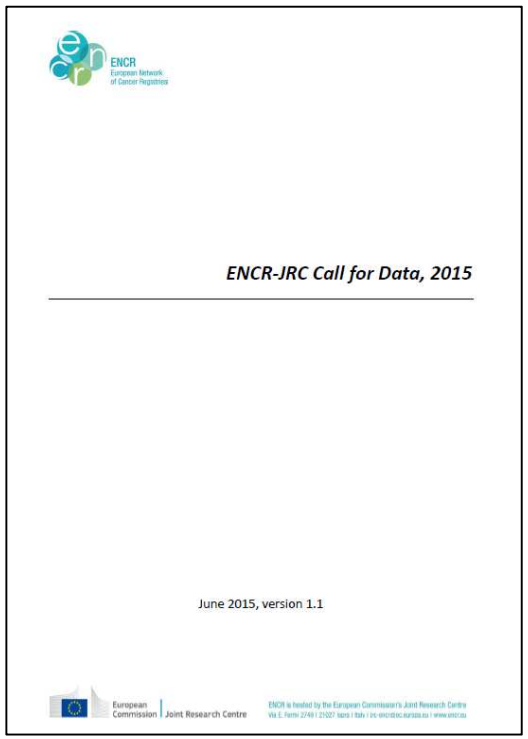
5 October 2016 - Baveno, Italy

Joint Research Centre

Reference documents



2014
Report on
cancer data quality checks



2015
Call for Data Protocol

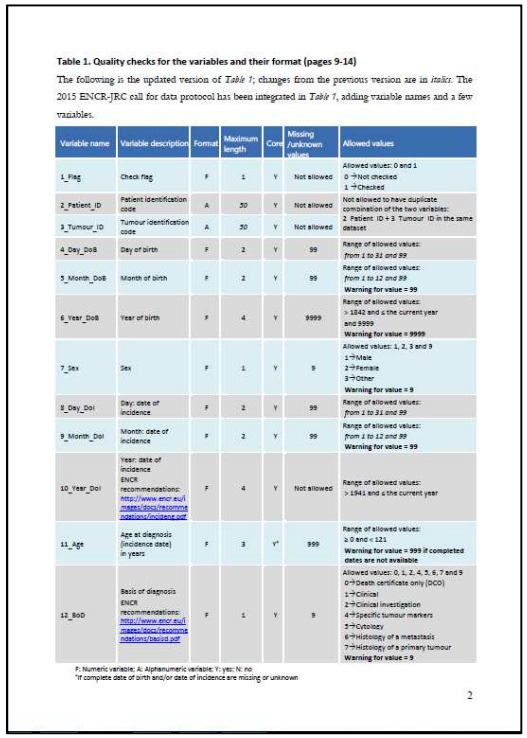


Table 1. Quality checks for the variables and their format (pages 9-14)

The following is the updated version of *Table 1*; changes from the previous version are in *italic*. The 2015 ENCR-JRC call for data protocol has been integrated in *Table 1*, adding variable names and a few variables.

Variable name	Variable description	Format	Maximum length	Core	Missing /unknown values	Allowed values
1_Flag	Check flag	F	1	Y	Not allowed	Allowed values: 0 and 1 0 →Not checked 1 →Checked
2_Patient_ID	Patient identification code	A	50	Y	Not allowed	Not allowed to have duplicate combination of the two variables: 2_Patient_ID + 3_Tumour_ID in the same dataset
3_Tumour_ID	Tumour identification code	A	50	Y	Not allowed	
4_Day_DoB	Day of birth	F	2	Y	99	Range of allowed values: *from 1 to 31 and 99*
5_Month_DoB	Month of birth	F	2	Y	99	Range of allowed values: *from 1 to 12 and 99* **Warning for value = 99**
6_Year_DoB	Year of birth	F	4	Y	9999	Range of allowed values: > 1842 and ≤ the current year and 9999 **Warning for value = 9999**
7_Sex	Sex	F	1	Y	9	Allowed values: 1, 2, 3 and 9 1 →Male 2 →Female 3 →Other **Warning for value = 9**
8_Day_DoI	Day: date of incidence	F	2	Y	99	Range of allowed values: *from 1 to 31 and 99*
9_Month_DoI	Month: date of incidence	F	2	Y	99	Range of allowed values: *from 1 to 12 and 99* **Warning for value = 99**
10_Year_DoI	Year: date of incidence ENCR recommendations: http://www.encr.eu/images/docs/recommendations/incidence.pdf	F	4	Y	Not allowed	Range of allowed values: > 1941 and ≤ the current year
11_Age	Age at diagnosis (incidence date) in years	F	3	Y*	999	Range of allowed values: ≥ 0 and < 121 **Warning for value = 999 if completed dates are not available**
12_BoD	Basis of diagnosis ENCR recommendations: http://www.encr.eu/images/docs/recommendations/basisd.pdf	F	1	Y	9	Allowed values: 0, 1, 2, 4, 5, 6, 7 and 9 0 →Death certificate only (DCO) 1 →Clinical 2 →Clinical investigation 4 →Specific tumour markers 5 →Cytology 6 →Histology of a metastasis 7 →Histology of a primary tumour **Warning for value = 9**

F: Numeric variable; A: Alphanumeric variable; Y: yes; N: no
*If complete date of birth and/or date of incidence are missing or unknown

2

2016
Addendum

Joint Research Centre

Checks overview

3.1. **Consistency within variables** ← **November 2015**

3.2. **Consistency between variables** ← **July 2016**

3.2.1. Coherence of dates

3.2.2. Consistency between tumour data and demographic information

- Consistency between age/topography/morphology
- Consistency between sex/topography
- Consistency between sex/morphology

+ Multiple primary tumours

3.2.3. Consistency between tumour variables

- Consistency between basis of diagnosis/morphology/behaviour
- Consistency between behaviour/topography/morphology
- Consistency between morphology/grade
- Consistency between topography/laterality
- Consistency between topography/morphology

3.3. **Specific additional checks for survival analysis** ← **July 2016**

3.4. **Other additional checks on the extent of the disease** ←

User ma

JRC TECHNICAL REPORTS

User manual for the JRC - ENCR Cancer Registries Data Quality Check Software (QCS)

QCS Version 1.7

September 2016

Francesco Giusti, Tadek Dyba, Carmen Martos, Raquel N. Carvalho, Nicholas Nicholson, Giorgia Randi, Roisin Rooney, Lydia Voti, Revveka Trigka, Stefano Adriani, Enrico Ben, Emanuele Crocetti, Manola Bettio

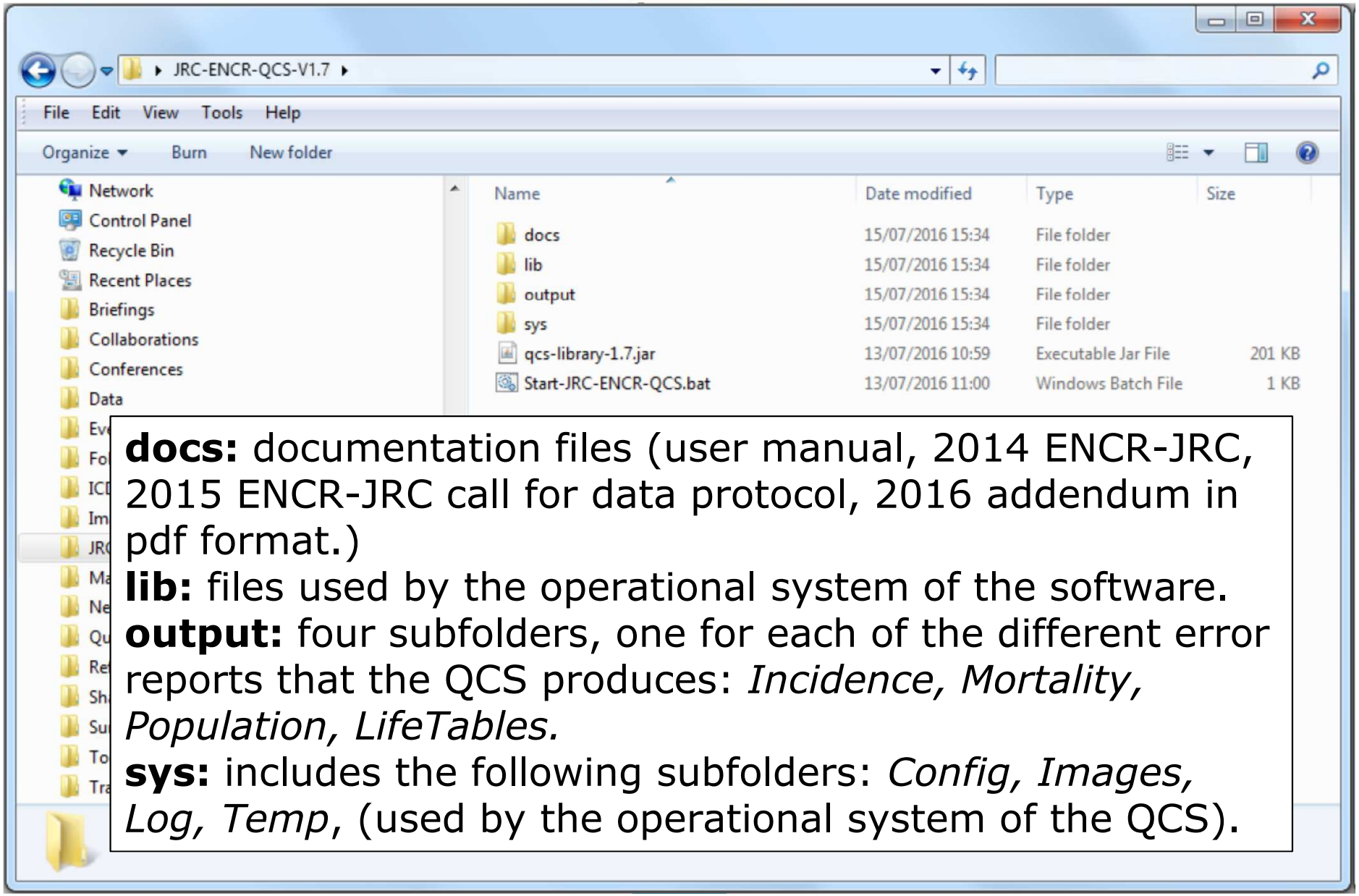
2016

Technical requirements

Minimum requirements:

- Windows Vista SP2 *OR* Windows 7 *OR* Windows 8 Desktop
- Support for DirectX 9 graphics and 32 MB of graphics memory
- Pentium 2 266 MHz or faster processor with at least 512 MB of physical RAM is recommended
- a minimum of 220 MB of free disk space to efficiently run the software

Java software is needed to run the JRC-ENCR-QCS!



docs: documentation files (user manual, 2014 ENCR-JRC, 2015 ENCR-JRC call for data protocol, 2016 addendum in pdf format.)

lib: files used by the operational system of the software.

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

sys: includes the following subfolders: *Config, Images, Log, Temp*, (used by the operational system of the QCS).

Analysis process

The software checks every single record for:

- number of variables
- variable names
- presence of non-missing values in the mandatory fields
- when applicable, the field content against a list of valid values
  ***Example**: patient's sex numeric value (variable 7_Sex) can be 1=male, 2=female, 3=other or 9=unknown. Every other value will produce an error*
- field length, which must be within the allowed range. ***Example**: maximum length for patient ID (variable 2_Patient_ID) is 50 characters*
- dates validity (also if dates are not set in the future)
- record failing the edits described in the 2014 ENCR-JRC Report "one common procedure for European cancer registries" (see also the related 2016 addendum)

Feedback received / enquiries about the QCS

- Estonia
- Finland
- France, Loire-Atlantique et Vendée
- Germany, Association of Population-based Cancer Registries (GEKID)
- Hungary
- Italy, Ragusa
- Italy, Varese
- Italy, Veneto
- Luxembourg
- Republic of Ireland
- Slovenia
- Spain, Basque Country
- Spain, Castellon
- Switzerland, Basel
- Switzerland, Bern
- Switzerland, National Institute for Cancer Epidemiology and Registration (NICER)
- Switzerland, St. Gallen-Appenzell