



European
Commission

Joint Research Centre

the European Commission's in-house science service



ENCR Scientific Meeting and General Assembly

From harmonization of quality check project to the development of the JRC-ENCR Quality Check Software

Carmen Martos
Health in Society Unit
JRC Directorate F – Health, Consumers and
Reference Materials

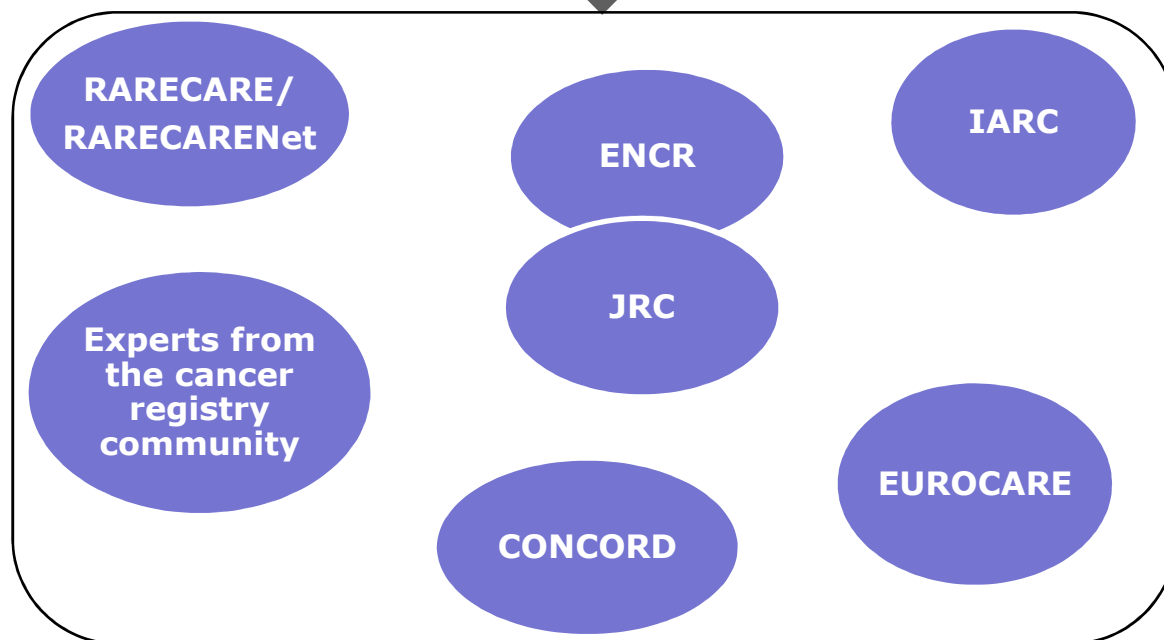


Quality check project

- ❑ The reliability and utility of information provided by population-based cancer registries (CRs) depends on the quality of the data collected.
- ❑ Several edits are in place to check the validity of European cancer registry data.
- ❑ The ENCR and the JRC launched in 2013 an initiative to establish a standardised list of cancer data quality checks to be adopted by European cancer registries in order to be used in European projects

Quality check project

Working Group



**3 Workshops
at JRC - Ispra (Italy)**

2 July 2013

Overview of data quality checks used by major EU studies and open discussion on quality control methods.

15 October 2013

Agreement was reached on the mandatory and basic variables list and their formats.

4 June 2014

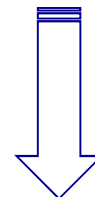
Discussion and agreement for "One common procedure for European Cancer Registries" 3

Quality check project



Focuses on

- Case definition
- Variables and their formats
- Internal consistency checks



Impossible codes or combinations → **Error**

Very unlikely codes or combinations → **Warning**

Not unlikely, but very rare code or combinations → **Warning**

Data quality

Validity

Completeness

Comparability

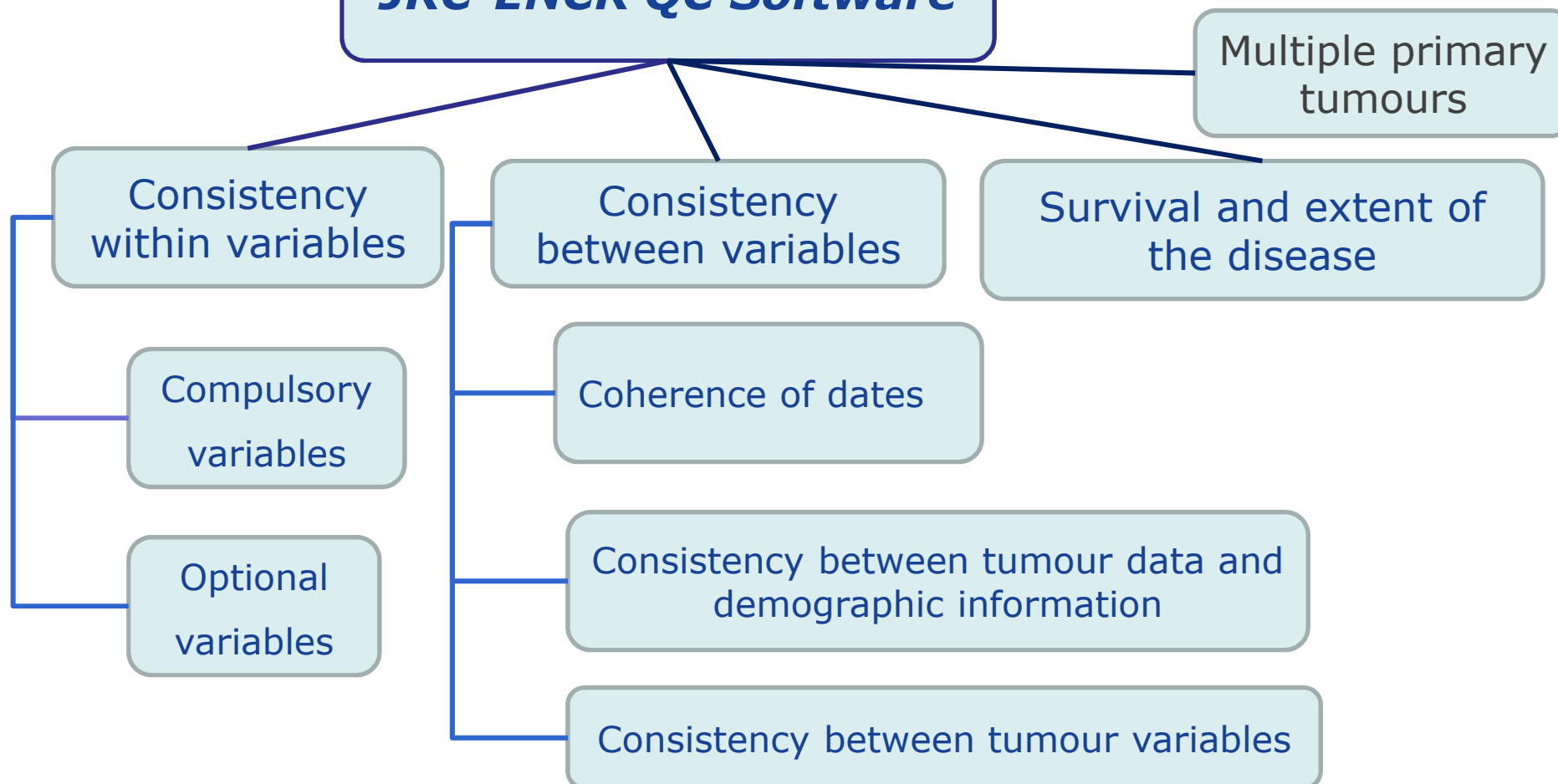
Timeliness

Internal
consistency



- ❑ The report provided the basis for the data quality-check software developed by the JRC.
- ❑ This toolkit is being used by the JRC for checking data submitted by European CRs for the 2015 ENCR-JRC data call.
- ❑ A new beta version of the software is available for downloading from the ENCR-JRC Portal.

JRC-ENCR QC Software



Consistency within variables

A proposal on cancer data quality checks: one common procedure for European cancer registries Addendum to version 1.0 (November 2014)

During the last two years the European Network of Cancer Registries (ENCR) and The Joint Research Centre (JRC) have been working in preparing the 2015 ENCR-JRC call for data and developing the JRC-ENCR Quality Check (QC) software.

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 *italics>*. 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, 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:

2_Pati 2						3_Tumo 1			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C619	8140	3	1	7/1/2009	3/4/1946	17_Autopsy	2	E-OUTR

2_Pati 8						3_Tumo 1			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C619	8140	3	a	7/1/2009	3/4/1946	7_Sex	a	E-FORM

Consistency within variables

A proposal on cancer data quality checks: one common procedure for European cancer registries Addendum to version 1.0 (November 2014)

During the last two years the European Network of Cancer Registries (ENCR) and The Joint Research Centre (JRC) have been working in preparing the 2015 ENCR-JRC call for data and developing the JRC-ENCR Quality Check (QC) software.

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 *italics>*. 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, 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_Dav_DoB	Dav of hirth	F	2	Y	99	Range of allowed values:

2_Pati 4						3_Tumo 1			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C619	8140	3	1	7/1/	3/4/1946	10_Year_DoI		E-MISS
							11_Age		E-MISS
							22_Survival		E-MISS

2_Pati 15						3_Tumo 1			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C910	8000	3	1	7/1/2009	3/4/1946	13_Topo	C910	E-OUTR

Consistency between tumour data and demographic information

Table 3. Unlikely and rare combinations of age and tumour type.

Pag.26-28

Age group [years]	Morphology	Topography
0-2	Hodgkin lymphoma: 9650-9667	-
> 9	Neuroblastoma and ganglioneuroblastoma: 9490, 9500	-
> 5	Retinoblastoma: 9510-9514	-

2_Pati 12						3_Tumo 1			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C779	9650	3	2	7/1/2009	3/4/2008	Calc_Age	0	W-AGMT
							14_Morpho	9650	W-AGMT
2_Pati 13						3_Tumo 2			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C504	8500	3	2	7/1/2009	3/4/1995	Calc_Age	13	W-AGMT
							13_Topo	C504	W-AGMT
							14_Morpho	8500	W-AGMT

Consistency between tumour data and demographic information

Table 5. Unlikely sex and morphology combinations.

Sex = 1 (male)	Sex = 2 (female)
8313/3 Clear cell adenocarcinofibroma	9061/3 Seminoma, NOS
8380/3 Endometrioid adenocarcinoma, NOS	9062/3 Seminoma, anaplastic
8381/3 Endometrioid adenofibroma, malignant	9063/3 Spermatocytic seminoma
8382/3 Endometrioid adenocarcinoma, secretory variant	
8383/3 Endometrioid adenocarcinoma, dilated cell variant	
8384/3 Adenocarcinoma, endocervical type	
8441/3 Serous cystadenocarcinoma, NOS	
8460/3 Papillary serous cystadenocarcinoma	
8471/3 Papillary mucinous cystadenocarcinoma	
8482/3 Mucinous adenocarcinoma, endocervical type	
8600/3 Thecoma, malignant	
8670/3 Steroid cell tumour, malignant	
8930/3 Endometrial stromal sarcoma, NOS	
8931/3 Endometrial stromal sarcoma, low grade	
8934/3 Carcinofibroma	
8950/3 Mullerian mixed tumour	
8951/3 Mesodermal mixed tumour	
9000/3 Brenner tumour, malignant	
9014/3 Serous adenocarcinofibroma	
9015/3 Mucinous adenocarcinofibroma	
9090/3 Struma ovarii, malignant	

Table 4. Invalid sex and topography combinations.

Sex = 1 (male)	Sex = 2 (female)
C51 Vulva	C60 Penis
C52 Vagina	C61 Prostate gland
C53 Cervix uteri	C62 Testis
C54 Corpus uteri	C63 Other and unspecified male genital organs
C55 Uterus NOS	

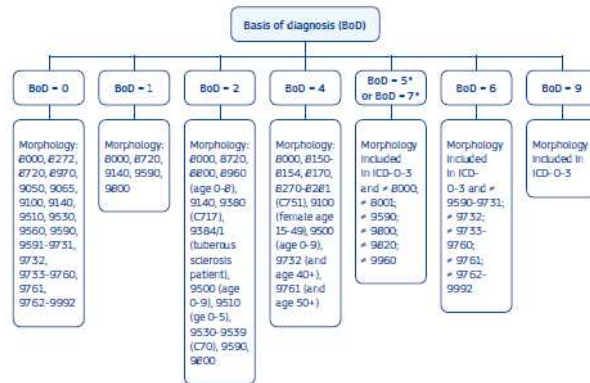
2_Pati 11		3_Tumo 1								
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code	
0	C619	8140	3	2	7/1/2009	3/4/1946	7_Sex	2	E-SETO	
							13_Topo	C619	E-SETO	

2_Pati 17		3_Tumo 1								
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code	
0	C619	8380	3	1	7/1/2009	3/4/1946	7_Sex	1	W-SEMO	
							14_Morpho	8380	W-SEMO	
							15_Beh	3	W-SEMO	

2_Pati 18		3_Tumo 2								
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code	
0	C539	9061	3	2	7/1/2009	3/4/1946	7_Sex	2	W-SEMO	
							14_Morpho	9061	W-SEMO	
							15_Beh	3	W-SEMO	



Figure 2. Valid combinations for basis of diagnosis and morphology.



Pag.30

Consistency between tumour variables

2_Pati 19						3_Tumo 1			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C509	8500	3	2	7/1/2009	3/4/1946	14_Morpho	8500	W-BDMO
							12_BoD	0	W-BDMO

2_Pati 20						3_Tumo 2			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C504	8000	3	1	7/1/2009	3/4/1946	14_Morpho	8000	W-BDMO
							12_BoD	7	W-BDMO

Table 7. Valid combinations for morphology and grade.

Grade →	5	6	8
Morphology	9700-9702, 9705, 9708, 9709, 9714, 9716, 9717, 9718, 9719, 9724, 9725, 9726, 9729, 9800, 9801, 9805-9807, 9809, 9820, 9827, 9831, 9834, 9837	9591, 9596, 9597, 9670, 9671, 9673, 9678-9680, 9684, 9687-9691, 9695, 9698, 9699, 9712, 9728, 9731, 9732, 9734, 9737, 9738, 9762, 9800, 9801, 9805-9808, 9811, 9812-9818, 9820, 9823, 9826, 9833, 9836, 9940	9719, 9727, 9831, 9948

Consistency between tumour variables

Pag.32

2_Pati 28		3_Tumo 1							
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C449	9700	3	2	7/1/2009	3/4/1946	16_Grade	6	W-MOGR
							14_Morpho	9700	W-MOGR
							15_Beh	3	W-MOGR

Pag.34-39

Table 9. Morphology codes and allowed/refused topography codes.

Morphology codes	Allowed topography codes	Not allowed topography codes
8000-8005		C420, C421, C77
8010-8589		C38, C40-C42, C47, C480, C49, C70-C72, C77
8015	C53	

Consistency between tumour variables

2_Pati 25						3_Tumo 1				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code	
0	C421	8000	3	1	7/1/2009	3/4/1946	14_Morpho	8000	W-MOTO	
							13_Topo	C421	W-MOTO	

2_Pati 26						3_Tumo 2				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code	
0	C189	8142	3	1	7/1/2009	3/4/1946	14_Morpho	8142	W-MOTO	
							13_Topo	C189	W-MOTO	

2_Pati 27						3_Tumo 3				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code	
0	C411	8720	3	1	7/1/2009	3/4/1946	14_Morpho	8720	W-MOTO	
							13_Topo	C411	W-MOTO	

3.3. Specific checks for survival analysis

Follow-up time and extent of disease are two important components to evaluate and interpret cancer survival.

- Consistency of vital status/autopsy/basis of diagnosis survival/dates of incidence

Vital status = 1 (alive)

Autopsy = 0 (not incidentally diagnosed at autopsy)

Basis of diagnosis ≠ 0 (DCO)

Pag.40-41

Survival and extent of the disease

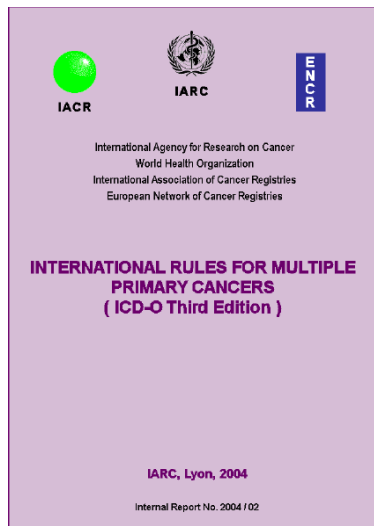
2_Pati 33					3_Tumo 4				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C504	8500	3	2	7/1/2009	3/4/1946	15_Beh	3	W-BTNM
							30_pT	is	W-BTNM

2_Pati 34					3_Tumo 5				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C559	8000	3	2	7/1/2009	3/4/1946	14_Morpho	8000	W-BDMO
							12_BoD	7	W-BDMO
							46_EoD	1	W-EoDT
							13_Topo	C559	W-EoDT

A proposal on cancer data quality checks:
one common procedure for European cancer registries
Addendum to version 1.0 (November 2014)

**Multiple primary
tumours**

During the last two years the European Network of Cancer Registries (ENCR) and The Joint Research Centre (JRC) have been working in preparing the 2015 ENCR-JRC call for data and developing the JRC-ENCR Quality Check (QC) software.

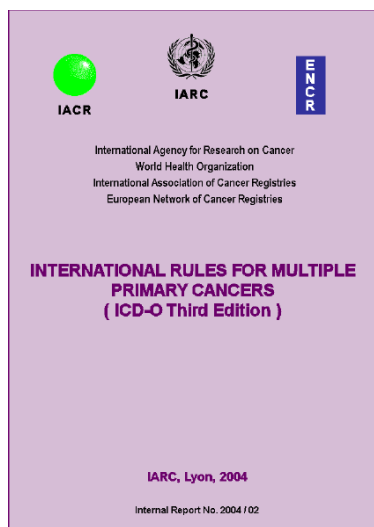


2_Pati 29						3_Tumo 1	
12_BoD	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Laterality
7	C504	8500	3	2	7/1/2009	3/4/1946	
2_Pati 29						3_Tumo 2	
12_BoD	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Laterality
7	C809	8500	3	2	12/5/2009	3/4/1946	

A proposal on cancer data quality checks: one common procedure for European cancer registries Addendum to version 1.0 (November 2014)

Multiple primary tumours

During the last two years the European Network of Cancer Registries (ENCR) and The Joint Research Centre (JRC) have been working in preparing the 2015 ENCR-JRC call for data and developing the JRC-ENCR Quality Check (QC) software.



2_Pati 30						3_Tumo 3	
12_BoD	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Laterality
7	C349	8070	3	1	7/3/2009	3/4/1946	
2_Pati 30						3_Tumo 4	
12_BoD	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Laterality
7	C349	8000	3	1	7/1/2009	3/4/1946	



Thank you for your attention

www.encr.eu

<https://ec.europa.eu/jrc>



JRC Science Hub:
ec.europa.eu/jrc



Twitter:
[@EU_ScienceHub](https://twitter.com/EU_ScienceHub)



YouTube:
[JRC Audiovisuals](https://www.youtube.com/JRC_Audiovisuals)



Facebook:
[EU Science Hub – Joint Research Centre](https://www.facebook.com/EU_Science_Hub_-_Joint_Research_Centre)



LinkedIn:
[Joint Research Centre \(JRC\) - European Commission's Science Service](https://www.linkedin.com/company/jrc-joint-research-centre)



Vimeo:
[Science@EC](https://vimeo.com/Science@EC)

Joint
Research
Centre

