

# European Network of Cancer Registries

**ENCR Workshop on Software and AI tools** for cancer registries

Łukasz Taraszkiewicz 18 October 2024, online

# Cancer cases data collection supported by the **eKRN+** application





# **Agenda**

- Introduction
- eKRN+
- Choosing standard dataset
- Updating and unifying
- Cancer notification form
- Dictionaries and validations
- Summary

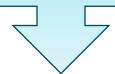




In Poland:

16 regional cancer registries

POPULATION-BASED



Polish National Cancer Registry (Warszawa)



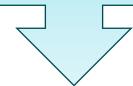




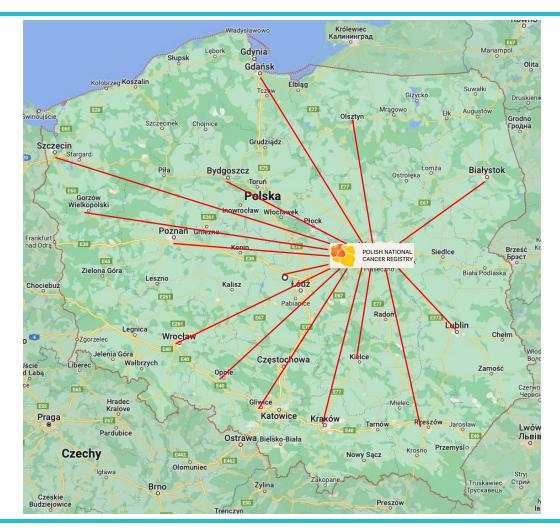
In Poland: 2023

16 regional cancer registries
PROH (site-specific)

POPULATION-BASED



Polish National Cancer Registry (Warszawa)





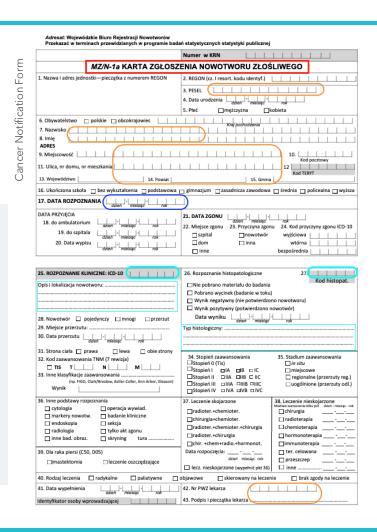


#### **PAPER ERA**

#### **Before 2013**

Minimum dataset required for a cancer case to be recorded in the registry (in colorful frames):

- ✓ Patient's ID number (PIN),
- ✓ Patient's Address,
- ✓ Tumour localisation,
- ✓ Tumour type,
- ✓ Professional license number.



#### INSTRUKCJA WYPEŁNIANIA KARTY ZGŁOSZENIA NOWOTWORU ZŁOŚLIWEGO

Zasady Ogólne

Karta "Głoszenia Nowotworu Złośliwego MZ/N-1a służy do zglaszania wykrytych przypadków nowotworów złośliwych oraz raka in situ przez placówki publicznej i niepublicznej służby zdrowia na terenie całego kraju.

Ratte należy wwoelniać:

przy pierwszym rozpoznaniu lub podejrzeniu nowotworu,

przy tych wizytach kontrolnych, w czasie których zostały stwierdzone istotne zmiany mające związek z:

diagnoza (zmiana lub doprecyzowanie rozpoznania, stwierdzenie kolejnego nowotworu); leczeniem (podjęcie i zakończenie leczenia, wdrożenie innego leczenia); postępem choroby (stwierdzenie przerzutów, nawrotu, progresji lub transformacji choroby).

na podstawie aktu zgonu, jeśli przyczyną zgonu był nowotwór złośliwy.

Zgłoszeniu podlegają chorby nowotworow oznaczone w X rewizji Międzynarodowej Statystycznej Klasyfikacji Chorób i Problemów Zdrowotnych numerami C00-C97, D00-D09 oraz D37-D48 (nowotwory o niepewnym lub nieznanym charakterze).

#### ZASADY WYPEŁNIANIA KARTY

Kartę należy wypełniać czytelnie czarnym lub niebieskim długopisem, używając drukowanych liter, a pola kodowe zaznaczać symbolem X wesunatrz obszani pola

Pola oznaczone szarym kolorem można pozostawić niewypełnione (nr 12, 25, 27).
Pole 1. Należy wstawić nazwę lub stempel jednostki zgłaszającej.

Pole 1. Naieży wstawie nazwę lub stempel jednostki zgraszającej.

Pole 2. Należy wpisać część I resortowego kodu identyfikacyjnego (REGON).

Pole 3. Należy wpisać pełny numer PESEL.

Pole 4. Należy wpisać datę urodzenia według ustalonego formatu: dd-mm-rrrr.

Pole 5. Należy zaznaczyć znakiem X odpowiednią pteć.
Pole 6. W przypadku obywateli innych krajów niż Polska, należy wpisać kraj pochodzeni

Pole 7. Należy wpisać aktualne nazwisko pacjenta.

Pole 8. Należy wpisać imię (imiona) pacjenta. Pole 9. Należy wpisać miejscowość, w której pacjent jest zameldowany na stałe

Pole 10. Należy wpisać kod pocztowy miejsca zamieszkania pacienta.

Pole 11. Należy wpisać ulicę, nr domu i mieszkania pacjenta w miejscu zameldowania

Pole 12. Nie wypełniać (wypełnia właściwy rejestr onkologiczny) Pole 13. Należy wpisać województwo, w którym leży miejscowoś

Pole 14. Należy wpisać powiat, w którym leży miejscowość

Pole 15. Należy wpisać gminę, w której leży miejscowość.

Pole 16. Należy zaznaczyć znakiem X ukończoną przez pacjenta szkolę.

Pole 17. Należy wpisać datę rozpoznania nowotworu według ustalonego formatu: dd-mm-rm

Pole 18. Należy wpisać datę porady ambulatoryjnej według ustalonego formatu: dd-mm-rrrr.

Pole 19. Należy wpisać datę przyjęcia do szpitala według ustalonego formatu: dd-mm-rrrr. Pole 20. Należy wpisać datę wypisu ze szpitala według ustalonego formatu: dd-mm-rrrr.

Pole 21. Należy wpisać datę zgonu pacjenta według ustalonego formatu: dd-mm-rrrr.

Pole 22. Należy zaznaczyć znakiem X miejsce zgonu pacjenta (hospicjum należy do kategorii "inne").

Pole 23. Należy zaznaczyć znakiem X przyczynę zgonu pacjenta.
Pole 24. Należy podać przyczynę zgonu w klasyfikacji ICD-10: wyjściową, wtórną i bezpośrednią.

Pole 25. Pola kodowego nie wypełniać. W miejscu na opis należy wpisać rozpoznanie kliniczne z dokładnym umiejscowieniem nowotworu (także tanki krwiotwórczej i układu chłomego), precyzując czy chodzi umiejscowienie pierwotne, czy przerzut, a pierwotny punkt wyjścia nowotworu jest nierzmy.

Pole 26. Jeżeli chory nie miał badania histopatologicznego, w polu należy zakreślić, "bie pobrano materiału do badania". Jeżeli nie otrzymano jeszcze swyniku badania, należy zakreślić, "pobrano wycinek - badanie w toku". Jeżeli wynik badania był ujenny, należy zakreślić "wynik padania był ujenny, należy zakreślić "wynik padania był ujenny, należy zakreślić "wynik pozytywny". W przypadku potwierdzenia nowotworu należy wpisać datę wyniku według ustalonego formatu: dd-mm-rrr orac (olownie) py histologiczny.

Pole 27. Pola kodowego nie wypełniać.

Pole 28. Należy zaznaczyć, czy rozpoznany nowotwór jest pierwszym nowotworem, czy jest to kolejny różny histologicznie nowotwór. W wypadku przerzutu zakreślić pole przerzut.

Pole 29. Wpisać (słownie) miejsce przerzutu nowotworu

Pole 30. Wpisać date wykrycia przerzutu.

Pole 31. W przypadku nowotworów umiejscowionych w narządach parzystych należy podać stronę ciała pacjenta, po której znajduje się

Pole 32. Należy wpisać kod zaawansowania choroby nowotworowej według klasyfikacji TNM wersja

Pole 33. Ješli istnije; psecyficzna dla noowtownu klasyfikacja zawansowania i znamy jest jej wynik, należy wpisia nazwe klasyfikacji, taktorej określono naślium zawawnowania cza wynik (pp. Masyfikacji pk. Odc.) a nowowów ginckologicznych (CSI-CSB), Ann. Arbor dla chłoniaków (CSI-CSB), Astler-Coller dla jelita grubego (CI8-C20), Bresłow/Clurk dla czerniaka (C43) lub sumę Gleasona dla raka gruzolu krokowego (CSI-CSB).

Pole 34. Należy określić stopień zaawansowania według klasyfikacji TNM wersja 7

Pole 35. Jeśli diagnostyka TNM nie jest możliwa, należy określić stadium zaawansowania choroby według podanych kategorii.

Pole 36. Jeżeli nie wykonano badania histopatologicznego, należy zakreślić, jakie inne badania były podstawą rozpoznania nowotworu. Wskazywanie innych badań, których rezultat nie posiada istotnej wartości diagnostycznej, jest niecelowe. W przypadku wykrycia nowotworu w badania ktyrningowym. nadeży rodać true badania. Możliwe ist zaznaczenie kilku nól.

Pole 37. Należy określić, jakie metody leczenia skojarzonego zastosowano u pacjenta (kolejność nieistotna) oraz datę wykonania badania.

Pole 38. Należy zakreślić, jakiemu leczeniu przeciwowotworowemu nieskojarzonemu poddano dotychczas pacjenta (operacja wywiadowcza nie jest leczeniem chirurgicznym). Przy każdoj metodzie należy wpisać datę rozpoczęcia leczenia. Możliwie jest zazaczenie killu pol. Pol 39. We rozywadku raka piersi (CS). DOS) podać, czy chirurgia obeimowała mastektomie czy leczenie oszczędziace (np. kwadzantektomia).

Pole 39. w przypadku raka piersi (C50, D05) podac, czy chirurgia obejmowata mastektomię czy iecze Pole 40. Należy określić rodzaj leczenia, zaznaczając jedną z przewidzianych w karcie możliwości.

Pole 41. Należy wpisać datę wypełnienia karty według formatu: dd-mm-rrrr

Pole 42. Numer PWZ lekarza wypełniającego kartę.

Pole 43. Czytelny podpis i pieczątka lekarza.



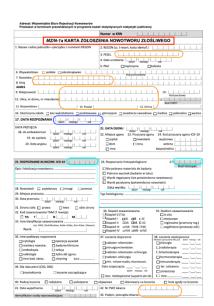


The first national web app designed for both registrars and healthcare professionals, featuring built-in validations.

# PAPER/ELECTRONIC

**ERA** 

**2013** 





Dane pacjenta Rozpoznanie		a: Dane pacjenta Rozpoznanie	Dane pacjenta:	Numer karty: 1 008 732 3
Dane jednostki 1. Nazwa jednostki	1A.Telefon do jed	Rozpoznanie kliniczne	Rozpoznanie histopatologiczne 26 Status badania wynik pozytywny (potwierdzono nowoty	Data wyniku wór) 2022-07-04
	5.Pleć 6.Kraj pochodzer <b>Mę</b> żczyzna	Opis i lokalizacja nowotworu Rak odbytnicy	27. Kod histopatologiczny 81403 - Gruzcolakorak BNO  Wynik histopatologiczny Rak gruczo?owy jelita - typ ?luzowy	
7.Nazwisko 8.Imię		25A. ICD-O-3 Kod topograficzny 27A.Kod histopograficzny 81403	atologiczny Stopień zróżnicowania	28. Nowotwór   Pojedynczy
PID		32. Kod zaawansowania TNM TS  T X - Guz pie N X - Region: M 1 - Obecno	34.Stopień 35.Stadium zaawansowania	29. Miejsce przerzutu 30. Data wykrycia
Dane adresowe pacjenta  9.Miejscowość  Kalisz  Bezdomny  10.Kod poczt  62-800	owy 11.Ulica	Klasyfikacja Astlera-Collera	37/38.Leczenie	31. Strona ciala
12.Województwo 13.Powiat WIELKOPOLSKIE Kalisz	14.Gmina Kalisz (gmina	36.Inne podstawy rozpoznania  miejsk Operacja wywiad.  Markery nowotw. Badanie kliniczne		Nie dotyczy Lewa Prawa Obie strony  40. Rodzaj leczenia
16.Ukończona szkoła		Endoskopia Sekcja Radiologia Tylko akt zgonu Inne badanie obrazowe Skrining  tura		
17.DATA ROZPOZNANIA 18.Do ambul 2022-06-17 Data przyjęcia:	19.Do szpitala 2024-01-04	20.Di 202		
21.Data zgonu 22.Miejsce zgonu	23.Przyczyna zgonu 24A.Wyjściowa	Komentarz		41.Data wypełnienia 42.Nr PWZ 2024-01-05 2
Ulicy nie ma w słowniku		Ulicy nie ma w słowniku		



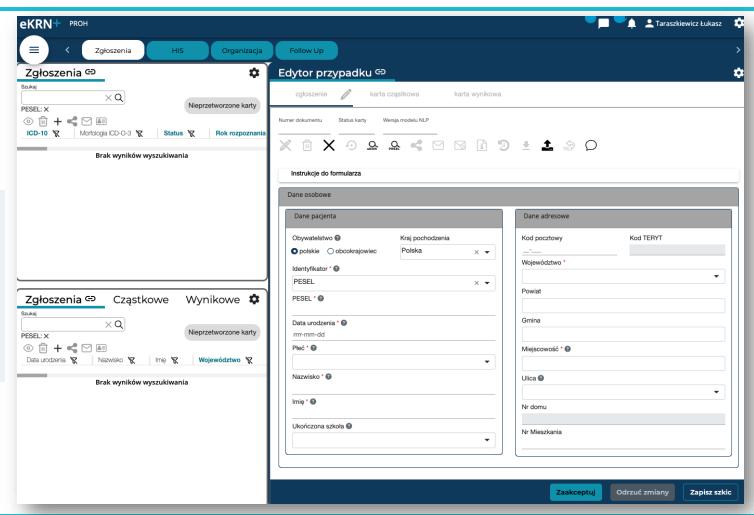




#### **ELECTRONIC ERA**

2023

The new e-KRN+ platform, established by the PNCR, is designed to collect and analyse cancer data with numerous built-in validations aligned with rules from JRC-ENCR Quality Check Software.

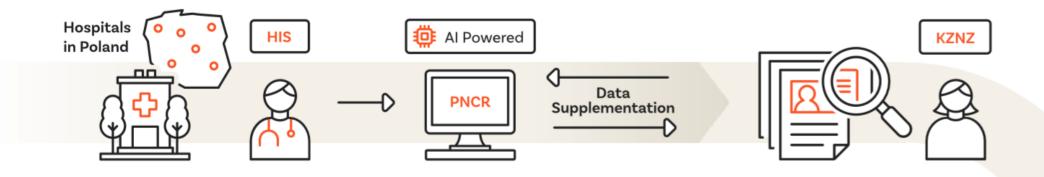








### e-KRN+



01

#### DATA SOURCES FOR PNCR

Data on new cancer cases can be reported to the Polish National Cancer Registry and site-specific registries both by physicians, though PNCR platform, and be directly downloaded from the PNCR-integrated hospital information systems (HIS).



#### **DATA VERIFICATION**

Reported/downloaded data is verified for its correctness and compliance with registration standards. Missing data are being filled in and the whole record registered or added to previously collected data.

Source: <a href="https://onkologia.org.pl/en">https://onkologia.org.pl/en</a>







### e-KRN+

#### AFTER VERIFICATION DATA IS TRANSFERRED TO CENTRAL DATABASE OF REGISTRIES



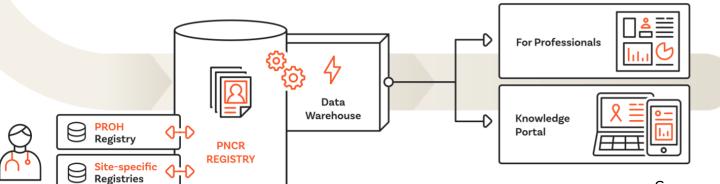
#### **CENTRAL DATABASE OF REGISTRIES**

Central Database of Registries is a technical environment, where all cancer registries are gathered (both PNCR and site-specific registries). Thanks to such solution data is not duplicated, but the common part of all registers is kept in one place. Due to the unification of forms, the person submitting data on the case simultaneously supplies several registers with a single notification.



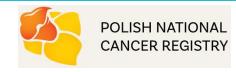
#### PORTAL, REPORTS, OPEN ACCESS DATA

The portal is a window on the world of the Polish National Cancer Registry, offering analyzes and reports not only for health professionals, but also presenting studies on cancer epidemiology at the national and voivodship level.



Source: <a href="https://onkologia.org.pl/en">https://onkologia.org.pl/en</a>



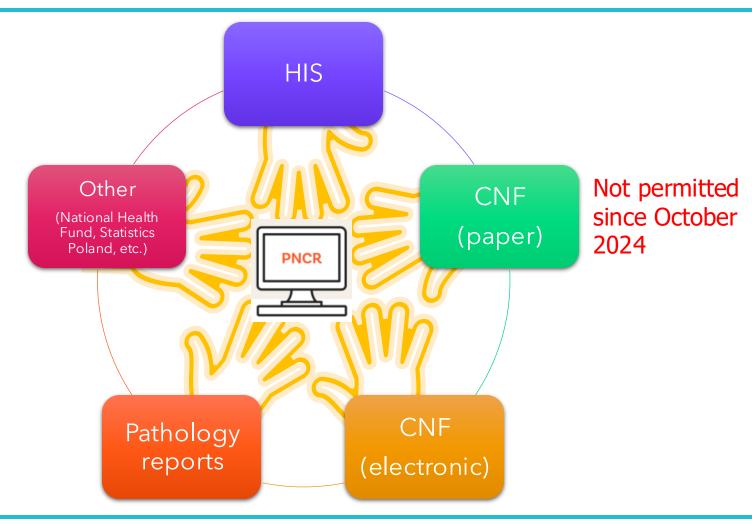




### e-KRN+ - data sources

#### Features of "input" data:

- Unstructured
- Not or hardly validated
- Incomplete
- InconsistentContradictoryOut of scope Contradictory









# Choosing standard dataset: rules/recommendations



Basis of diagnosis



Standard dataset

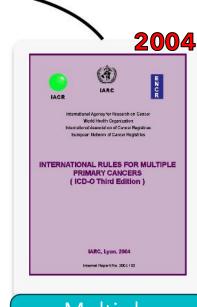


tumours





entities

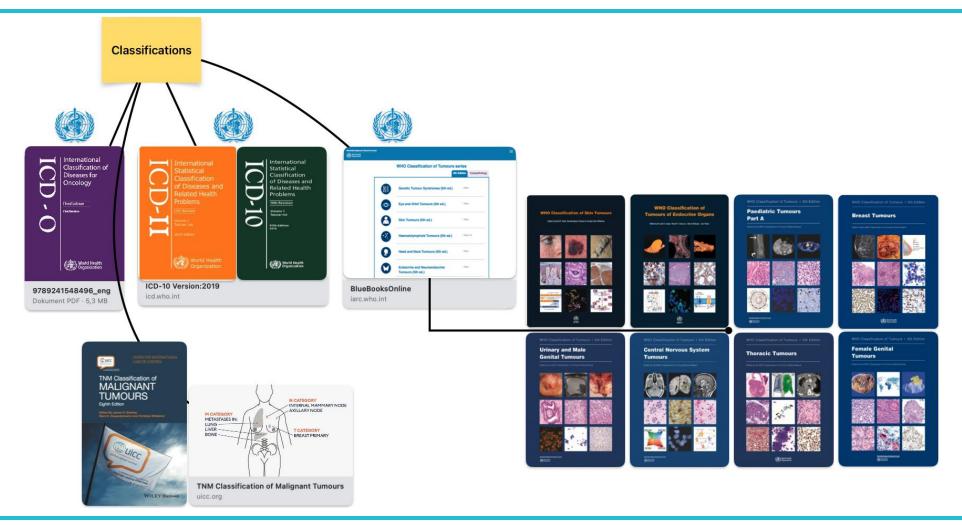


Multiple primaries





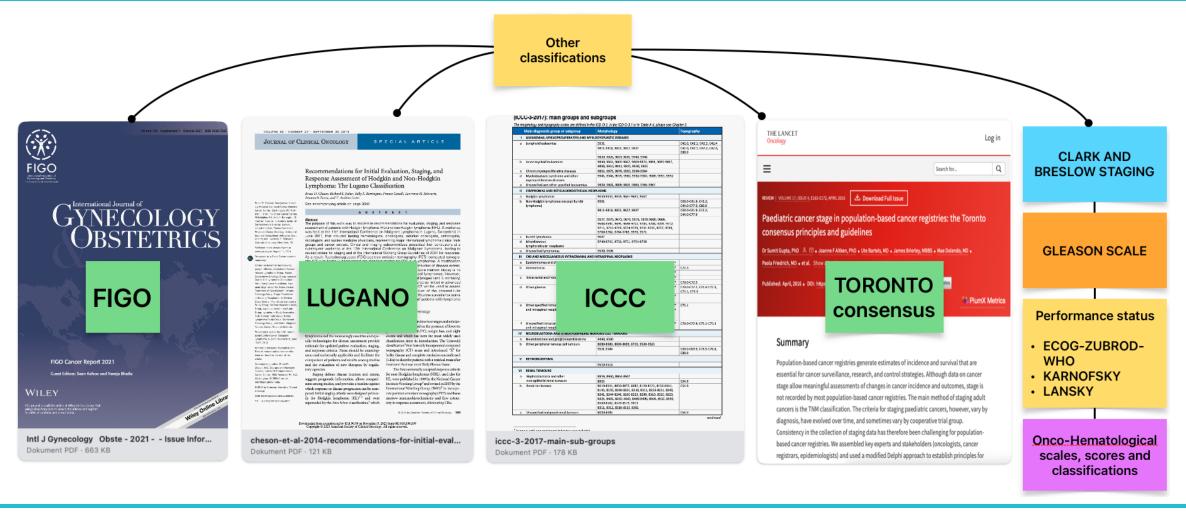
# **Choosing standard dataset: classifications**







# **Choosing standard dataset: classifications**







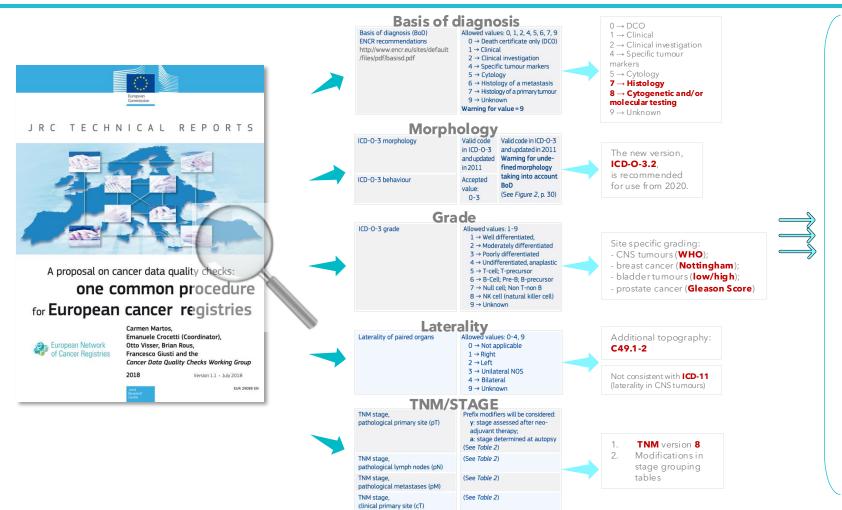
# Helpers







# **Quality Check Rules - Updating and unifying**





"Consistency between variables" needed to be updated and harmonized (with **ICD-O3** and "**Blue Books**").

Consistency between age/topography/morphology

Consistency between sex/topography

Consistency between sex/morphology

Consistency between basis of diagnosis/morphology/behaviour

Consistency between morphology/grade

Consistency between topography/laterality

Consistency between topography/morphology



**"LITERALITY"** had to be included in the algorithm of recognizing multiple primary tumours.



"TNM 8 edition" tables had to be created

Table 2. Valid values for T, N, M and stage by cancer topography and TNM edition.<sup>5</sup>

Topography	TNM edition	Т	N	М	Stage grouping
Lip and oral cavity COO, CO2-CO6	6	TX, T1, T2, T3, T4a, T4b	NX, N0, N1, N2, N2a, N2b, N2c, N3	MX, M0, M1	I, II, III, IVA, IVB, IVC
(except C051 and C052)	7	TX, T1, T2, T3, T4a, T4b	NX, N0, N1, N2, N2a, N2b, N2c, N3	M0, M1	I, II, III, IVA, IVB, IVC

Appendix III: TNM 7 edition stage grouping and corresponding T, N, M values

Lip and Oral Cavity C00, C02-C06 (except C051 and C052)								
Stage	Т	N	М					
0	Tis	NO	мо					
1	Т1	NO	МО					
Ш	T2	NO	МО					
Ш	T1,T2,T3	N1	МО					
	T3	NO	MO					

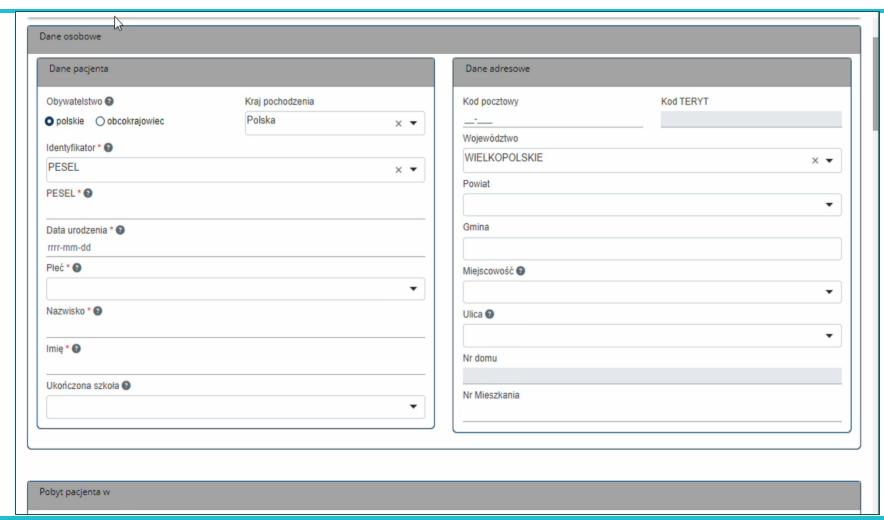




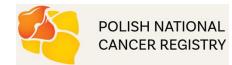
2022



### **Cancer notification form**

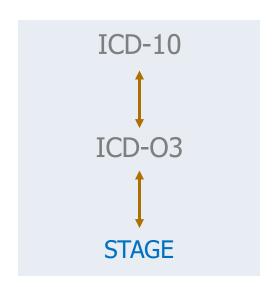


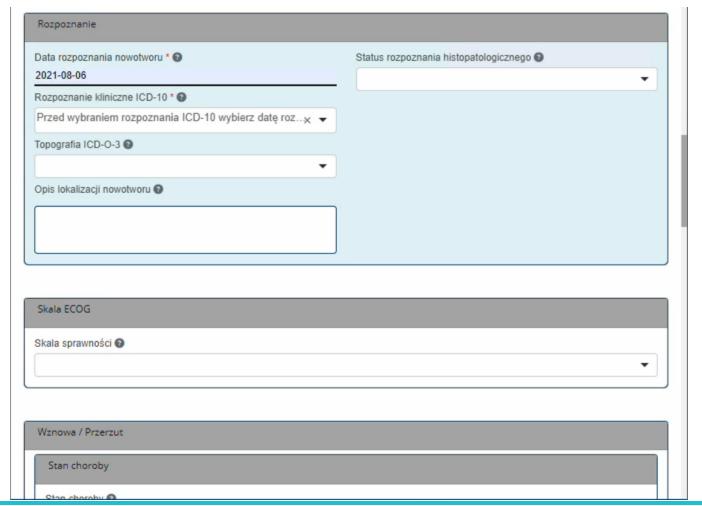




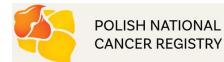


# **Cancer notification form – consistency: STAGE**



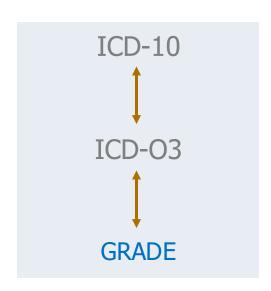


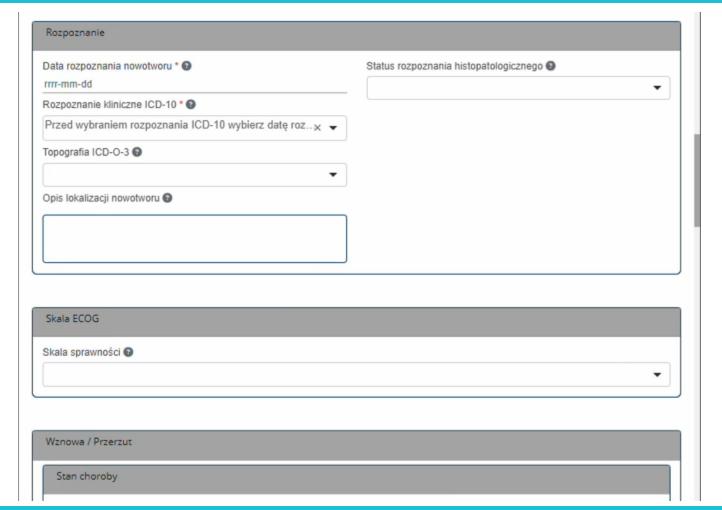




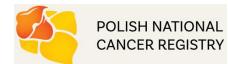


# **Cancer notification form – consistency: GRADE**



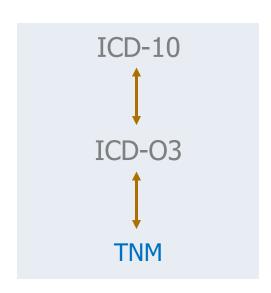


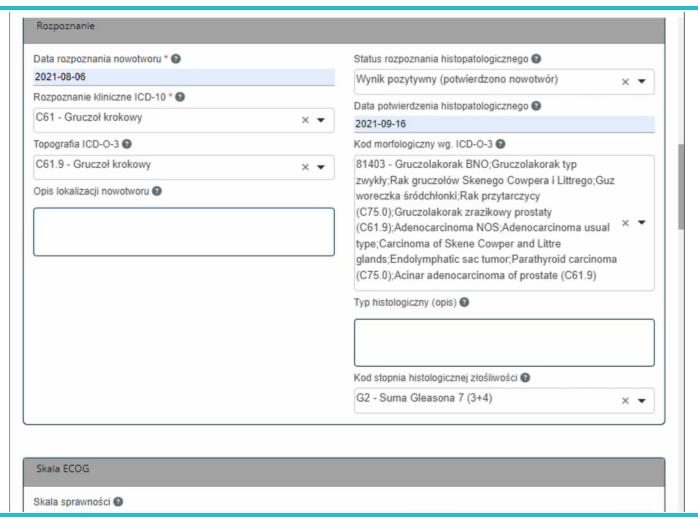




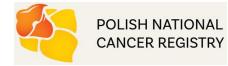


# Cancer notification form – consistency: TNM



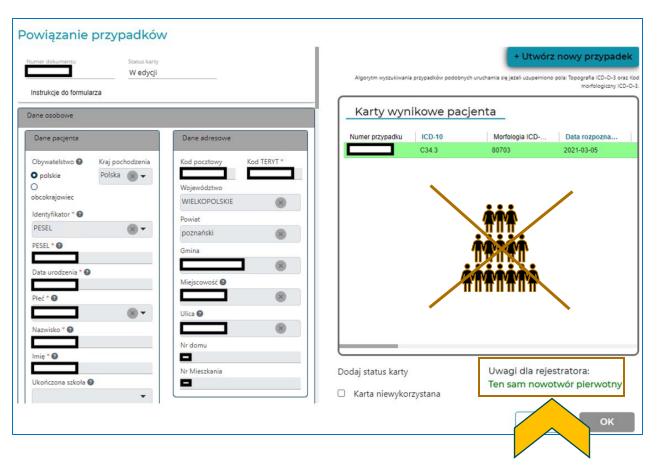




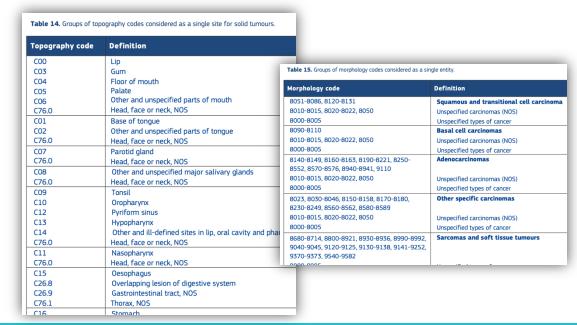




# **Multiple primaries**







Multiple primaries checking tool







# Dictionaries and validations – QCS version



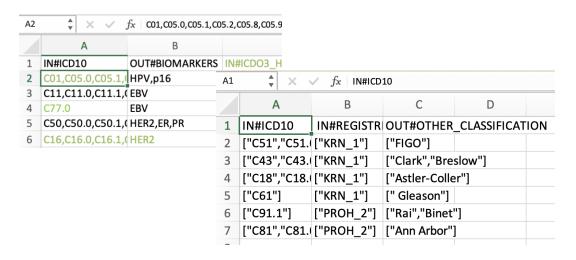
encr2014	∷≡≎	<b>₩ ~</b> ① ◇	⊕ ×	Q
	Data zmian	Wielkość	Rodzaj	
ozwala przeglądać poprzednio oglądane foldery.  AgeTumourRule.csv	8.03.2020, 18:35	2 KB	Dokument CS	
BehaviourMorphologyRule.csv	11.01.2021, 11:26	10 KB	W- V	
BehaviourMorphologyTopographyRule1.csv	8.03.2020, 18:35		Dokument CS	
BehaviourMorphologyTopographyRule2.csv	8.03.2020, 18:35	154 B		
BehaviourMorphologyTopographyRule3.csv	8.03.2020, 18:35		Dokument CS	
BehaviourMorphologyTopographyRule4.csv	8.03.2020, 18:35		Dokument CS	
			Dokument CS	
BehaviourMorphologyTopographyRule5.csv	8.03.2020, 18:35		Dokument CS	
DiagnosisMorphologyRule1.csv DiagnosisMorphologyRule2.csv	8.03.2020, 18:35 8.03.2020, 18:35		Dokument CS	
	to the same of the			
DiagnosisMorphologyRule3.csv	8.03.2020, 18:35	29 B		
DiagnosisMorphologyRule4.csv	8.03.2020, 18:35	34 B	725	I NMRuleEultiono.yami
DiagnosisMorphologyRule5.csv	8.03.2020, 18:35	29 B		TNM Edition 8
DiseaseExtentRule1.csv	8.03.2020, 18:35	174 B	#	INIT EDICTION O
DiseaseExtentRule6.csv	8.03.2020, 18:35	114 B	A SAN TENNES OF THE SAN THE SA	
DiseaseExtentRule9.csv	8.03.2020, 18:35	39 B		ımber: 8
EditionCauseOfDeathRule1.csv	8.03.2020, 18:35	41 KB	me	orphologies:
Grade.csv	8.03.2020, 18:35	17 B	Dokument (	_ id: M1
GradeMorphologyRule1.csv	8.03.2020, 18:35		Dokument (	description: M1
GradeMorphologyRule2.csv	8.03.2020, 18:35	197 B	8	data: 000-8005,8010-8012,8014-8015,8020-8022,8030-8035,8050,8051-8081,8083,8084,8090-8098,8046,8082,8102-8110,8
ICDEdition.csv	8.03.2020, 18:35	11 B		3131,8140-8149,8170-8175,8160-8163,8180,8190-8231,8250-8384,8390-8420,8430,8440-8552,8570-8576,8560-8562,8
Morphology.csv	8.03.2020, 18:35	4 KB	Dokument (8-	-8589,8940-8941,9014-9015,9110,8950-8951,8980-8983,9000,9390
MorphologyTopographyRule1.csv	8.03.2020, 18:35	30 KB	Dokument (	<pre>- id: M1a description: M1a</pre>
MorphologyTopographyRule2.csv	8.03.2020, 18:35		Dokument (	data: '
PrimaryMultipleContext_1.csv	11.01.2021, 11:46	19 KB		000-8005,8010-8012,8014-8015,8020-8022,8030-8035,8050,8090-8098,8046,8082,8102-8110,8140-8149,8170-8175,8
PrimaryMultipleContext_2.csv	11.01.2021, 11:46	5 KB		3163,8180,8190-8231,8250-8384,8390-8420,8430,8440-8552,8570-8576,8560-8562,8588-8589,8940-8941,9014-9015,9
PrimaryMultipleCriteria.csv	28.05.2020, 13:32	213 B	Dokument (	- id: M1b
SexMorphologyRule1.csv	8.03.2020, 18:35	104 B	Dokument (	description: M1b
SexMorphologyRule2.csv	8.03.2020, 18:35	14 B	Dokument C	data: 8051-8081,8083,8084,8120-8131 - id: M2
SexTopographyRule1.csv	8.03.2020, 18:35	134 B	Dokument (	description: M2
SexTopographyRule2.csv	8.03.2020, 18:35	74 B	Dokument (	data: 8720-8790
TopographyLateralityRule.csv	8.03.2020. 18:35	264 B	Dokument (	<pre>- id: M3 description: M3 data: 8050,8260,8341,8343,8340,8052,8130,8342,8344,8350,8450-8452,8290,8330-8332,8335</pre>

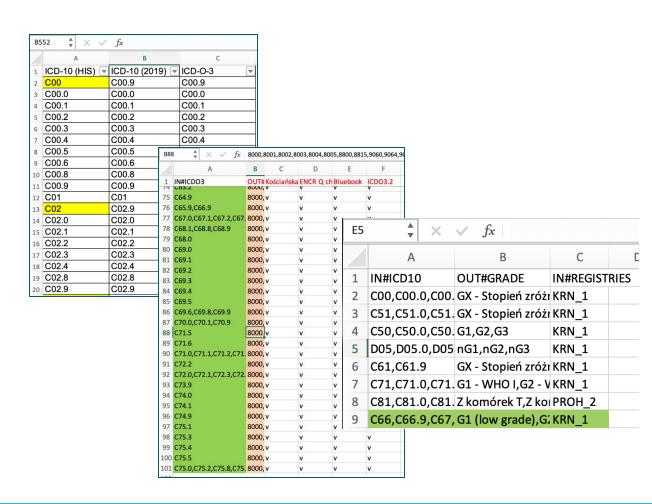




### Dictionaries and validations — tailored version

	Α	В	С	D	E	F	G	Н	1	J	K	L
1	IN#REV_TN	ICD10_TOPO	IN#AGE	IN#MITOTIC_RA	IN#GRADE	IN#Seru	IN#RISK_CA	IN#MORPHOLOGY	IN#cT	▼ IN#cN	IN#cM	OUT#STAGE
03	REV8	C52.9	-					8010,8011,8020,8	T1, T2, T3	N1	M0	III
04	REV8	C52.9	-					8010,8011,8020,8	T4	#All values#	M0	IVA
05	REV8	C52.9	-					8010,8011,8020,8	#All values#	#All values#	M1	IVB
06	REV8	C53.0, C53.1, C53.8, C53.9	-					8010,8011,8013,8	Tis	N0	M0	0
07	REV8	C53.0, C53.1, C53.8, C53.9	-					8010,8011,8013,8	T1	N0	M0	1
80	REV8	C53.0, C53.1, C53.8, C53.9	-					8010,8011,8013,8	T1a	N0	M0	IA
09	REV8	C53.0, C53.1, C53.8, C53.9	-					8010,8011,8013,8	T1a1	N0	MO	IA1
10	REV8	C53.0, C53.1, C53.8, C53.9	-					8010,8011,8013,8	T1a2	N0	M0	IA2
11	REV8	C53.0, C53.1, C53.8, C53.9	-					8010,8011,8013,8	T1b	N0	MO	IB
12	REV8	C53.0, C53.1, C53.8, C53.9	-					8010,8011,8013,8	T1b1	N0	M0	IB1
13	REV8	C53.0, C53.1, C53.8, C53.9	-					8010,8011,8013,8	T1b2	N0	MO	IB2
14	REV8	C53.0, C53.1, C53.8, C53.9	-					8010,8011,8013,8	T2	N0	M0	II
15	REV8	C53.0, C53.1, C53.8, C53.9	-					8010,8011,8013,8	T2a	N0	MO	IIA
16	REV8	C53.0, C53.1, C53.8, C53.9	-					8010,8011,8013,8	T2a1	N0	M0	IIA1
17	REV8	C53.0, C53.1, C53.8, C53.9	-					8010,8011,8013,8	T2a2	N0	MO	IIA2

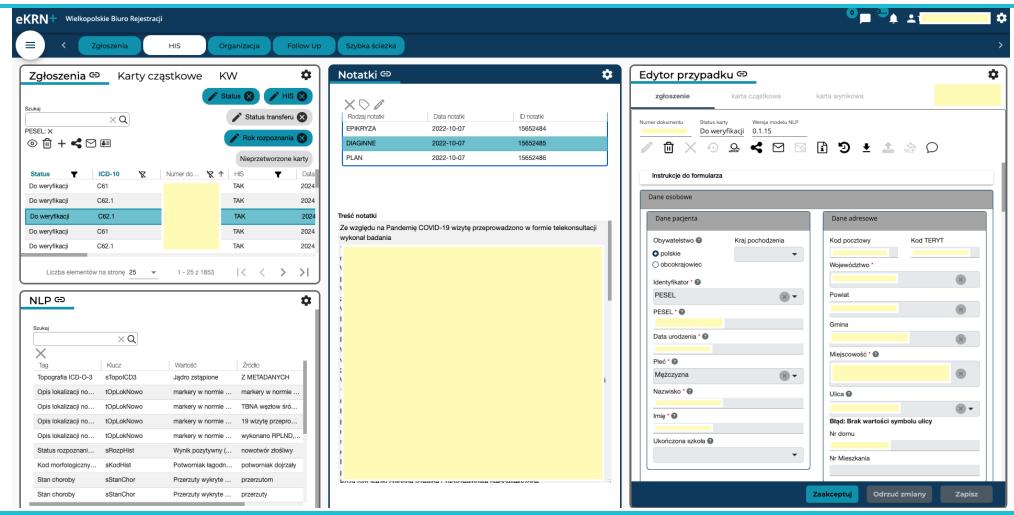








# NLP – the most interesting part of the system









# **Summary**

- Creating a registry system/application to record and report cancer data requires the combined efforts of both registrars and IT specialists.
- Quality Check Software is a complex IT tool that is extremely helpful in creating validations in registry systems.
- The need to update its rules and classifications when implementing validations in cancer registration systems is both crucial and challenging.
- Validation should help registrars avoid making obvious mistakes, but it shouldn't prevent them from coding unlikely but real combinations of variables.





# "e-KRN+" project team

Urszula Wojciechowska, team leader

Joanna Didkowska

Maciej Trojanowski

Irmina Michałek

Łukasz Taraszkiewicz

Urszula Sulkowska

Paweł Olasek

Beata Kościańska

Klaudia Barańska

Marta Miklewska

Agnieszka Jawołowska



# Let's talk...

# Thank you!



