

Bundesweiter klinischer Krebsregisterdatensatz - Datenschema und Klassifikationen

Robert Koch-Institut | RKI

Nordufer 20
13353 Berlin

Zentrum für Krebsregisterdaten | ZfKD

Stefan Meisegeier (Projektleitung)

Maren Imhoff (Data Manager)

Karsten Berg (Data Analyst)

Klaus Kraywinkel (Leitung ZfKD)

E-Mail-Adresse für Rückmeldungen: krebsdaten@rki.de

Zitieren

Meisegeier, S., Imhoff, M., Berg, K. und Kraywinkel, K. (2023): Bundesweiter klinischer Krebsregisterdatensatz - Datenschema und Klassifikationen (oBDS_v3.0.0.8a_RKI). Zenodo. DOI:[10.5281/zenodo.10022040](https://doi.org/10.5281/zenodo.10022040)

Einleitung

Die Krebsregistrierung in Deutschland erfolgt auf der Basis von Landesgesetzen. Diese verpflichten medizinische Einrichtungen (v. a. niedergelassene Ärztinnen und Ärzte, pathologische Institute, Kliniken, Screening-Einheiten), neu auftretende Krebsfälle und definierte Ereignisse im Krankheits- bzw. Behandlungsverlauf an das zuständige Krebsregister zu melden.

Die Krebsregister der Bundesländer wiederum übermitteln nach Vorgabe des [Bundeskrebsregisterdatengesetzes \(BKRG\)](#) einmal jährlich Angaben zu neu erfassten Erkrankungsfällen an das Zentrum für Krebsregisterdaten (ZfKD) am Robert Koch-Institut. Das ZfKD prüft die Qualität der Daten, führt sie zu einem bundesweiten Datensatz zusammen und stellt sie [auf Antrag für wissenschaftliche Forschungsprojekte](#) zur Verfügung.

In diesem Repository werden begleitende Informationen zu [Struktur](#) und [Klassifikationen](#) des bundesweiten ZfKD-Datensatzes bereitgestellt.

Die hier verwendeten [Klassifikationen](#) spiegeln den derzeitigen Arbeitsstand des ZfKD wider. Ein wesentliches Ziel des Repositories ist es, diesen Stand möglichst transparent abzubilden und beteiligte Akteure zur weiteren gemeinsamen Harmonisierung von Standards einzuladen.

Der ZfKD-Datensatz ist nicht öffentlich zugänglich, kann aber auf Antrag für wissenschaftliche Forschungszwecke genutzt werden. Bitte verwenden Sie für Fragen zur Antragstellung die oben genannte E-Mail-Adresse oder das auf der Internetseite des ZfKD bereitgestellte [Kontaktformular](#). Informationen zum gesetzlichen Auftrag, zu Methoden und Veröffentlichungen des ZfKD erhalten Sie ebenfalls auf den [Internetseiten des ZfKD](#). Bitte beachten Sie, dass das ZfKD an den Daten, die von den Krebsregistern übermittelt wurden, keine Änderungen vornimmt.

Informationen zum Entstehungskontext des ZfKD-Datensatzes

Für die Erhebung klinischer Krebsregisterdaten wurde mit dem [Krebsfrüherkennungs- und -registergesetz \(KFRG\)](#) im [§ 65c Fünftes Buch Sozialgesetzbuch \(SGB V\)](#) ein bundesrechtlicher Rahmen geschaffen. Die von den klinischen Krebsregistern zu erfassenden Angaben werden in dem von der Arbeitsgemeinschaft Deutscher Tumorzentren (ADT) und der Gesellschaft der epidemiologischen Krebsregister in Deutschland (GEKID) erarbeiteten [onkologischen Basisdatensatz \(oBDS\)](#) spezifiziert und regelmäßig überarbeitet. Die letzte Anpassung des oBDS wurde am 12. Juli 2021 [im Bundesanzeiger publiziert](#). Einmal jährlich übermitteln die Krebsregister Daten nach Maßgabe des [Bundeskrebsregisterdatengesetzes \(BKRG\)](#) an das ZfKD.

Seit der Novellierung des BKRG durch das [Gesetz zur Zusammenführung von Krebsregisterdaten](#) enthalten die ans ZfKD übermittelten Daten auch klinische Angaben, u. a. zum Krankheitsverlauf und zur Behandlung (ab Diagnosejahr 2020).

Die Inhalte und die Struktur der ans ZfKD zu übermittelnden Daten wurden in einer AG mit Vertretern des ZfKD und der Krebsregister abgestimmt, dabei diente der oBDS und das novellierte Bundeskrebregisterdatengesetz (§5) als Arbeitsgrundlage.

Das Arbeitsergebnis ist das hier beschriebene, für die Datenübermittlung ans ZfKD zu verwendende XML-Schema (alternativ als oBDS-RKI oder ZfKD-Lieferdatensatz bezeichnet, siehe dazu [Struktur des bundesweiten klinischen Krebsregisterdatensatzes](#)).

Umfassende Informationen zur Krebsregistrierung sind hier verfügbar: [Manual der klinischen und epidemiologischen Krebsregistrierung](#) (Veröffentlichung 2019)

Administrative und organisatorische Angaben

Das [Zentrum für Krebsregisterdaten \(ZfKD\)](#) des RKI ist zuständig für die bundesweite Krebsberichterstattung und stellt Dritten auf Antrag Daten für überregionale Forschungsprojekte zur Verfügung. Es prüft die Qualität der von den Krebsregistern übermittelten Daten und gibt den Krebsregistern diesbezüglich Rückmeldung.

Inhaltliche Fragen zur Datenerhebung, Datenauswertung und Datenkuration können direkt an das ZfKD gestellt werden (E-Mail-Adresse für Anfragen: krebsdaten@rki.de).

Datenübermittlung an das ZfKD

Das 2009 verabschiedete BKRK regelt die jährliche Zusammenführung der wesentlichen Daten aus den Krebsregistern am ZfKD. Die Übermittlung erfolgt jeweils am Jahresende und enthält Informationen zu allen Fällen, die bis zum Ende des vorherigen Kalenderjahres diagnostiziert wurden, so dass auch Nachmeldungen und Korrekturen sowie Informationen zum Follow-up (z. B. Sterbefälle und Wegzüge) früherer Erkrankungsfälle enthalten sind.

Vor der Novellierung des BKRK in 2021 wurde lediglich der deutlich kleinere epidemiologische Datensatz (mit Angaben zur Diagnose und zum Sterbezeitpunkt) an das ZfKD übermittelt. Dieser Datensatz wird bundesweit seit 2009 erfasst. Die Mehrzahl der Bundesländer hat zwischen 1998 und 2007 mit der landesweiten Erfassung begonnen.

Seit der Datenlieferung zum 31. Dezember 2022 und rückwirkend ab dem Diagnosejahr 2020 liefern die Krebsregister [auch klinische Angaben](#). Die am ZfKD vorliegenden Daten enthalten allerdings nicht den gesamten Datenbestand der Register, beispielsweise sind keine Angaben zu den behandelnden Einrichtungen verfügbar.

Außerdem sind die Daten in den Krebsregistern bearbeitet worden: So wurden Meldungen aus verschiedenen Quellen zum gleichen Erkrankungsfall zusammengeführt und weitgehend um Widersprüche bereinigt („best-of“). Der Datensatz des ZfKD ist daher fall- und nicht meldungsbasiert, mehrere Tumorerkrankungen derselben Person können anhand einer von den Registern einmal vergebenen Personidentifikationsnummer zugeordnet werden. Die Übermittlung der Daten an das ZfKD erfolgt nach dem Wohnortprinzip (zum Zeitpunkt der Diagnose), so dass Doppelmeldungen weitgehend ausgeschlossen sind. Zwischen den Bundesländern erfolgt ein regelmäßiger Austausch von Daten, die außerhalb des Wohnortbundeslandes der Erkrankten erhoben und zunächst an das Krebsregister des Behandlungsortes gemeldet wurden.

Eine fallweise Verknüpfung (Record Linkage) der am ZfKD vorliegenden Daten mit externen Datensätzen (Studien, Krankenkassen) ist nicht möglich.

Struktur des bundesweiten klinischen Krebsregisterdatensatzes

Der klinische Datensatz wird als **oBDS-RKI** bezeichnet. Die Bezeichnung geht zurück auf den zwischen ADT, GEKID und Plattform § 65c abgestimmten **einheitlichen onkologischen Basisdatensatz (oBDS)**, der für die Entwicklung des **oBDS-RKI** als Vorlage und Arbeitsgrundlage diente (siehe [Informationen zum Datensatz und Entstehungskontext](#)).

Weil er die Struktur und Inhalte der von den Landeskrebsregistern ans ZfKD zu liefernden Daten definiert, wird der **oBDS-RKI** auch als **ZfKD-Lieferdatensatz** bezeichnet.

Zur Veranschaulichung der Datenstruktur werden zufällig generierte [Beispieldaten](#) bereitgestellt.

Folgende Informationen sind enthalten:

- **Datenschema** des Datensatzes in verschiedenen Formaten
- **Klassifikationen**: Referenztabelle für Variablen des Datensatzes und ihre definierten Ausprägungen
- **Beispieldaten**: zur Veranschaulichung des Bereitstellungsprozesses der Daten

Datenschema

Das Datenschema umfasst mehr als 120 Variablen, die verschiedenen Elementen zugeordnet sind. Die klinischen Daten können nicht in einer einfachen „Rechtecktafel“ wiedergegeben werden, da sie zum Teil komplexe Krankheitsverläufe abbilden. Im klinischen Datensatz sind die Daten daher in einem verschachtelten XML-Schema strukturiert.

Der klinische Datensatz wird durch folgende Elemente gegliedert:

- Die *Person* bildet die grundlegende Einheit im Datensatz.
- Der *Person* zugeordnet ist mindestens ein Element *Tumor*.
- Das Element *Tumor* enthält ein verpflichtendes Element *Primärdiagnose*. Dieses enthält u. a. Angaben zum Tumorstadium, zur Histologie und Lokalisation des Tumors.
- Darüber hinaus sind dem Element *Tumor* mehrere optionale Elemente zugeordnet, in denen Angaben zur Behandlung (Elemente *OP*, *ST* und *SYST*) und zu Folgeereignissen (Element *Folgeereignis*) wie Remissionen und Rezidiven erfasst werden können.

Bestimmte Variablen sind Pflichtangaben, z. B. das *Geburtsdatum*, der *Inzidenzort* und der *Diagnoseschlüssel*. Viele Angaben sind optional, z. B. die den Elementen *cTNM* und *pTNM* zugeordneten Variablen (*T-Kategorie*, *UICC-Stadium*, *m-Suffix* usw.). Einige Angaben sind nur unter der Bedingung verpflichtend, dass das übergeordnete, optionale Element verwendet wird: Beispielsweise ist das Element *Histologie* optional. Wird jedoch in der zugehörigen Variable *Morphologie* ein Eintrag vorgenommen, ist auch eine Angabe zum *Grading* verpflichtend. Angaben zur Zahl untersuchter Lymphknoten bleiben optional.

Bei Auswertungen ist zu beachten, dass optionale Inhalte möglicherweise nicht gleichermaßen aus allen Bundesländern vorliegen.

Die Elemente *Primärdiagnose*, *Folgeereignis*, *OP*, *ST* und *SYST* können mehrfach verwendet werden, so dass auch komplexe Krankheitsverläufe abgebildet werden können. Die Inhalte eines Elements können in ein tabellarisches Format überführt und über eine fallbezogene Nummer mit anderen Tabellen aus dem Datensatz verknüpft werden. Auf diese Weise entsteht ein auswertbares Format, in dem die bewilligten Daten an den Datenempfänger übermittelt werden können (siehe [Beispieldaten](#)).

Protokollierte Änderungen am Datenschema sind in den beigefügten [Release Notes](#) der Versionen zu finden.

Abbildung: Vereinfachtes Datenschema (mit ausgewählten Variablen). Quelle: krebsdaten.de.

Downloads

Das Datenschema wird in verschiedenen Formaten zum Download angeboten:

Datei	Beschreibung	Download
XML-Schema	Die XML-Schema-Definition <code>.xsd</code> als eindeutige, vollständige und maschinenlesbare Repräsentation des gesamten Schemas mit allen Details.	
CSV-Schema	Variablen und mögliche Ausprägungen in tabellarischer Darstellung als <code>.csv</code> .	
TXT-Schema	Variablen und mögliche Ausprägungen in stark vereinfachter textueller Darstellung zur erleichterten Erkennung von Änderungen.	
PDF-Schema (Abbildung)	Die grafische Darstellung des XML-Schemas als <code>.pdf</code> . Aufgrund der Komplexität des Gesamtschemas sind nicht alle Elemente abgebildet. Hinweise zur Notation des XML-Schemas sind hier zu finden.	
PDF-Schema (Liste)	Optisch gestaltete und "druckerfreundliche" Kurzübersicht zu Variablen und möglichen Ausprägungen als <code>.pdf</code> .	

XML-Schema des Datensatzes

Eine vollständige und maschinenlesbare Repräsentation des gesamten Datenschemas mit allen Details ist über das [XML-Schema](#) bereitgestellt.

XML (Extensible Markup Language)-Schemata definieren den erlaubten Aufbau der ihnen zugeordneten XML-Dokumente. XML ist eine Auszeichnungssprache mit definierter Struktur und Syntax. XML-Dokumente sind textbasiert und repräsentieren Daten in einer hierarchischen und strukturierten Weise. Der Hauptzweck von XML besteht darin, Daten so zu beschreiben, dass sie sowohl für Menschen als auch für Maschinen leicht verständlich und interpretierbar sind.

Ein XML-Schema, oft auch als XSD (XML Schema Definition) bezeichnet, bietet einen Rahmen zur Beschreibung der Struktur und Datentypen eines XML-Dokuments. XML-Schemata legen fest, welche Elemente und Attribute in einem XML-Dokument erscheinen können, wie diese strukturiert und organisiert sind und welche Datentypen sie enthalten können. XML-Schemata können dazu verwendet werden, um XML-Dokumente zu validieren. Hierbei wird überprüft, ob ein XML-Dokument der im Schema definierten Struktur entspricht.

Detaillierte technische Informationen zum abgestimmten XML-Schema sind auf der [Internetseite der Plattform § 65c abrufbar](#) (bis Version 3.0.0.8_RKI).



Abbildung: Übersicht zum XML-Schema des klinischen Datensatzes. Quelle: eigene Darstellung.

Klassifikationen

Die für einzelne Variablen erwarteten Ausprägungen und ihre Beschreibung sind in [Referenztabellen](#) hinterlegt. Einzelne Referenzen werden für mehrere Variablen genutzt: Beispielsweise wird für den Östrogen-Rezeptorstatus und den Progesteron-Rezeptorstatus die gleiche Kodierung verwendet. Ebenso werden für die Angaben zur klinischen und pathologischen TNM die gleichen Referenztabellen genutzt.

Größtenteils handelt es sich bei den Referenzen um Vereinbarungen, die bei der Erarbeitung des ZfKD-Lieferdatensatzes getroffen wurden (z. B. Ausprägungen von Variablen im Element Strahlentherapie, Ausprägungen von *Diagnosesicherung*). Teilweise handelt es sich bei den Referenzen um internationale oder nationale Standards (z. B. TNM, ATC-Klassifikation für den deutschen Arzneimittelmarkt). Informationen zu Quelle und Version der jeweiligen Referenzwerte, zu ihrer Interpretation und zu gegebenenfalls bestehenden Nutzungsbedingungen der Herausgeber sind im Abschnitt [Ergänzungen zu den Referenztabellen](#) zusammengestellt.

Einige Referenztabellen geben Inhalte von Standards wieder, die von Dritten herausgegeben werden. Unter Umständen verbinden diese Anbieter die Nutzung ihrer Produkte mit Bedingungen. Die Nutzungsbedingungen sind an den jeweils zutreffenden Stellen verlinkt. Wir bitten Sie diese zu beachten.

Referenztabellen

In der folgenden Übersicht sind die verwendeten Referenztabellen aufgeführt. Die Tabellen bilden den Wertebereich aller *kodierten* Variablen als Klassifikationen ab.

Bedeutung der Spalten:

- **Klassifikation** ausformulierte Bezeichnung der Klassifikation
- **Element** technische Variablenbezeichnung Elternknoten im XML-Schema sowie der technische Name der entsprechenden Variable. Dieser Name wird u.a. auch im exportierten Datensatz verwendet.
- **Datei** Name der Klassifikationsdatei

Klassifikation	Element technische Variablenbezeichnung	Da
----------------	---	----

Angabe zur perkutanen Strahlentherapie	<i>Strahlentherapie</i> Atemgetriggert	atemgetrigg
Folgeereignis - Gesamtbeurteilung Tumorstatus	<i>Folgeereignis</i> Gesamtbeurteilung_Tumorstatus	beurteilung_
Folgeereignis - Beurteilung Primärtumor	<i>Folgeereignis</i> Verlauf_Lokaler_Tumorstatus	beurteilung_
Wertigkeit der Diagnosesicherung	<i>Primärdiagnose</i> Diagnosesicherung	diagnosesic
Lokalisation der Fernmetastasen	<i>Primärdiagnose</i> Lokalisation , <i>Folgeereignis</i> Lokalisation	fm_lokalisat
Geschlecht	<i>Person</i> Geschlecht	geschlecht.c
Modul Prostata: Anlass der Probenahme	<i>Primärdiagnose</i> AnlassGleasonScore	gleason_anl
Modul Prostata: Gleason-Score	<i>Primärdiagnose</i> ScoreErgebnis	gleason_sco
Differenzierungsgrad	<i>Primärdiagnose</i> Grading	grading.csv
Modul Mamma: Hormonrezeptorstatus	<i>Primärdiagnose</i> HormonrezeptorStatus_0estrogen , <i>Primärdiagnose</i> HormonrezeptorStatus_Progesteron	hormonreze
Todesursache, Grundleiden nach ICD-10	<i>Todesursachen</i> Code	icd10_todes
Ausgabe der ICD-10	<i>Todesursachen</i> Version , <i>Primärdiagnose</i> Diagnose_ICD10_Version	icd10_versic
Diagnose nach ICD-10	<i>Primärdiagnose</i> Diagnose_ICD10_Code	icd10.csv
Angabe zur Kontaktbestrahlung	<i>Strahlentherapie</i> Interstitiell_endokavitaer	interstitiell.c
Wohnort bei Diagnose	<i>Primärdiagnose</i> Inzidenzort , <i>Primärdiagnose</i> Inzidenzort_BL	landkreis.cs
Modul Mamma: Menopausenstatus	<i>Primärdiagnose</i> Praetherapeutischer_Menopausenstatus	menopause
Typ der metabolischen Strahlentherapie	<i>Strahlentherapie</i> Metabolisch_Typ	metabolisch

Quelle Morphologie	<i>Primärdiagnose</i> Morphologie_Version	morphologie
Morphologie	<i>Primärdiagnose</i> Morphologie_Code	morphologie
Intention der OP	<i>Operation</i> Intention	op_intention
Operationen- und Prozedurenschlüssel (OPS)	<i>Operation</i> Code	ops.csv
Therapieprotokoll	<i>Systemische Therapie</i> Protokoll_TypProtokollschluessel_Code	protokoll.csv
Ausführung der perkutanen Radiochemotherapie	<i>Strahlentherapie</i> Radiochemo	radiochemo
Modul Darm: Mutation K-ras-Onkogen	<i>Primärdiagnose</i> RASMutation	rasmutation
Dosisleistung Kontaktbestrahlung	<i>Strahlentherapie</i> Rate_Type	rate_type.csv
Körperseite der bestrahlten Region	<i>Strahlentherapie</i> Seite_Zielgebiet	seite_zielgebiet
Seitenlokalisierung bei paarigen Organen	<i>Primärdiagnose</i> Seitenlokalisierung	seitenlokalisierung
Intention der Strahlentherapie	<i>Strahlentherapie</i> Intention	st_intention.csv
Bezug Strahlentherapie - OP	<i>Strahlentherapie</i> Stellung_OP	st_op_stellung
Angabe zur perkutanen Strahlentherapie	<i>Strahlentherapie</i> Stereotaktisch	stereotaktisch
Verwendete Substanzen	<i>Systemische Therapie</i> Type0fSYST_TypSubstanz	substanz.csv
Intention der systemischen Therapie	<i>Systemische Therapie</i> Intention	syst_intention.csv
Bezug systemische Therapie - OP	<i>Systemische Therapie</i> Stellung_OP	syst_op_stellung
Art der systemischen Therapie	<i>Systemische Therapie</i> Therapieart	therapieart.csv
TNM-Ausgabe	<i>Primärdiagnose</i> TNM_Auflage_c , <i>Primärdiagnose</i> TNM_Auflage_p , <i>Folgeereignis</i> Version	tnm_auflage

TNM-Präfix (c, p, u)	<i>Primärdiagnose</i> c_p_u_Praefix_T_c , <i>Primärdiagnose</i> c_p_u_Praefix_N_c , <i>Primärdiagnose</i> c_p_u_Praefix_M_c , <i>Primärdiagnose</i> c_p_u_Praefix_T_p , <i>Primärdiagnose</i> c_p_u_Praefix_N_p , <i>Primärdiagnose</i> c_p_u_Praefix_M_p , <i>Folgeereignis</i> c_p_u_Praefix_T , <i>Folgeereignis</i> c_p_u_Praefix_N , <i>Folgeereignis</i> c_p_u_Praefix_M	tnm_cpu.csv
TNM: Lymphgefäßinvasion	<i>Primärdiagnose</i> cTNM L_p , <i>Primärdiagnose</i> pTNM L_p , <i>Folgeereignis</i> L	tnm_l.csv
TNM: Fernmetastasierung	<i>Primärdiagnose</i> cTNM M_c , <i>Primärdiagnose</i> pTNM M_p , <i>Folgeereignis</i> M	tnm_m.csv
TNM: Regionäre Lymphknotenmetastasierung	<i>Primärdiagnose</i> cTNM N_c , <i>Primärdiagnose</i> pTNM N_p , <i>Folgeereignis</i> N	tnm_n.csv
TNM: Perineuralinvasion	<i>Primärdiagnose</i> cTNM Pn_c , <i>Primärdiagnose</i> pTNM Pn_p , <i>Folgeereignis</i> Pn	tnm_pn.csv
TNM: Serumtumormarker	<i>Primärdiagnose</i> cTNM S_c , <i>Primärdiagnose</i> pTNM S_p , <i>Folgeereignis</i> S	tnm_s.csv
TNM: Ausdehnung des Primärtumors	<i>Primärdiagnose</i> cTNM T_c , <i>Primärdiagnose</i> pTNM T_p , <i>Folgeereignis</i> T	tnm_t.csv
TNM: UICC-Stadium	<i>Primärdiagnose</i> cTNM UICC_Stadium_c , <i>Primärdiagnose</i> pTNM UICC_Stadium_p , <i>Folgeereignis</i> UICC_Stadium	tnm_uicc.csv
TNM: Veneninvasion	<i>Primärdiagnose</i> cTNM V_c , <i>Primärdiagnose</i> pTNM V_p , <i>Folgeereignis</i> V	tnm_v.csv
Ausgabe der ICD-O	<i>Primärdiagnose</i> Topographie_Version	topographie
ICD-O Topographie	<i>Primärdiagnose</i> Topographie_Code	topographie
Verlauf: Fernmetastasierung	<i>Folgeereignis</i> Verlauf_Tumorstatus_Fernmetastasen	verlauf_fern

Verlauf: Lokaler Tumorstatus	<i>Folgeereignis</i> Verlauf_Lokaler_Tumorstatus	verlauf_loka
Verlauf: Regionärer Lymphknotenstatus	<i>Folgeereignis</i> Verlauf_Tumorstatus_Lymphknoten	verlauf_lym
Zielgebiet Strahlentherapie oBDS2014	<i>Strahlentherapie Perkutan</i> CodeVersion2014 , <i>Strahlentherapie Kontakt</i> CodeVersion2014 , <i>Strahlentherapie Metabolisch</i> CodeVersion2014 , <i>Strahlentherapie Sonstige</i> CodeVersion2014 , <i>Strahlentherapie Unbekannt</i> CodeVersion2014	zielgebiet_2
Zielgebiet Strahlentherapie oBDS2021	<i>Strahlentherapie Perkutan</i> CodeVersion2021 , <i>Strahlentherapie Kontakt</i> CodeVersion2021 , <i>Strahlentherapie Metabolisch</i> CodeVersion2021 , <i>Strahlentherapie Sonstige</i> CodeVersion2021 , <i>Strahlentherapie Unbekannt</i> CodeVersion2021	zielgebiet_2

Datumsangaben

Die Angabe *Tag* wird von den Registern grundsätzlich *nicht* ans ZfKD übermittelt. Das ZfKD legt den Tag auf einen beliebigen Wert fest (i. d. R. 15), so dass immer ein Datum im Format *Jahr-Monat-Tag* vorliegt. Für jede Datumsangabe im Datensatz liegen jeweils zwei Variablen vor:

- das *Datum* im internationalen Datumsformat (ISO 8601) *yyyy-mm-dd* und
- die *Genauigkeit* des Datums in einer von drei möglichen Ausprägungen (M , T , V):
 - M = nur das Jahr ist bekannt (jahrgenau)
 - T = Jahr und Monat sind bekannt (monatsgenau)
 - V = Jahr und Monat wurden geschätzt

Ergänzungen zu den Referenztabellen

In diesem Abschnitt werden ergänzende Informationen zu den Inhalten der [Referenztabellen](#) bereitgestellt.

Diagnose nach ICD-10

Die Erstellung der Referenztablette erfolgte unter Verwendung der maschinenlesbaren Fassung der [ICD-10-GM \(Version 2008\)](#) des Bundesinstituts für Arzneimittel und Medizinprodukte (BfArM), unter Verwendung der [Empfehlungen des GKV-Spitzenverbands für die klinische Krebsregistrierung \(Stand: 14.05.2020\)](#) und unter Verwendung des [Umsetzungsleitfadens der Plattform § 65c \(Stand: 06.06.2023\)](#).

Die Nutzungsbedingungen der ICD-10 des BfArM sind [hier](#) hinterlegt.

[Klassifikationen/icd10.csv](#)

Variablen und Ausprägungen der Referenztablette:

Variable	Typ	Ausprägungen	Beschreibung
id	String	z. B. C021	ICD-10-Diagnoseschlüssel, 4-Steller werden ohne Trennzeichen dargestellt
code	String	z. B. C02.1	ICD-10-Diagnoseschlüssel, 4-Steller werden mit Trennzeichen dargestellt
name	String	z. B. Bösartige Neubildung...	Beschreibung der Diagnose
id3	String	z. B. C02	ICD-10-Diagnoseschlüssel, 3-stellig
epi_valide	Boolean	TRUE , FALSE	Information, ob die Diagnose im epidemiologischen Datensatz des ZfKD enthalten ist
p65_valide	Boolean	TRUE , FALSE	Es besteht eine Meldepflicht für den klinischen Datensatz (lt. Plattform § 65c-Umsetzungsleitfaden).

Todesursache, Grundleiden nach ICD-10

Die Erstellung der Referenztablette erfolgte unter Verwendung der maschinenlesbaren Fassung der [ICD-10-GM \(Version 2022\)](#) des Bundesinstituts für Arzneimittel und Medizinprodukte (BfArM). Die Nutzungsbedingungen der ICD-10 des BfArM sind [hier](#) hinterlegt.

Um die internationale Vergleichbarkeit zu gewährleisten, ist für die [Verschlüsselung von Todesursachen die ICD-10-WHO](#) vorgesehen. Aktuell wird bei der Übermittlung von Todesursachen ans ZfKD vorwiegend (noch) die Verwendung der ICD-10-GM angegeben.

[Klassifikationen/icd10_todesursache.csv](#)

Variablen und Ausprägungen der Referenztablette:

Variable	Typ	Ausprägungen	Beschreibung
id	String	z. B. C021	ICD-10-Diagnoseschlüssel, 4-Steller werden ohne Trennzeichen dargestellt
code	String	z. B. C02.1	ICD-10-Diagnoseschlüssel, 4-Steller werden mit Trennzeichen dargestellt
name	String	z. B. Bösartige Neubildung...	Beschreibung der Diagnose
id3	String	z. B. C02	ICD-10-Diagnoseschlüssel, 3-stellig
chapter	Integer	z. B. 1	ICD-10-Kapitelnummer

ICD-O Topographie

Die Erstellung der Referenztablelle erfolgte unter Verwendung der maschinenlesbaren Fassung der [ICD-O-3 \(2. Revision, Version 2019\) des Bundesinstituts für Arzneimittel und Medizinprodukte \(BfArM\)](#) und unter Zuhilfenahme des [Umsetzungsleitfadens der Plattform § 65c \(Stand: 10.05.2023\)](#).

Die Nutzungsbedingungen der ICD-O-3 des BfArM sind [hier](#) hinterlegt.

Für paarige Organe (Ausprägung *istPaarig* = 1, lt. Plattform § 65c-Umsetzungsleitfaden) wird bei der Variable *Seitenlokalisierung* die Angabe der betroffenen Körperseite(n) erwartet.

[Klassifikationen/topographie.csv](#)

Variablen und Ausprägungen der Referenztablelle:

Variable	Typ	Ausprägungen	Beschreibung
id	String	z. B. C021	ICD-10 Diagnoseschlüssel, 4-Steller werden ohne Trennzeichen dargestellt
code	String	z. B. C02.1	ICD-10 Diagnoseschlüssel, 4-Steller werden mit Trennzeichen dargestellt
name	String	z. B. Bösartige Neubildung...	Beschreibung der Diagnose
id3	String	z. B. C02	3-stelliger Diagnoseschlüssel
istPaarig	Integer	z. B. 1	1 = es handelt sich um ein paariges Organ und es wird bei der Variable <i>Seitenlokalisierung</i> die Angabe der betroffenen Körperseite(n) erwartet.

Operationen- und Prozedurenschlüssel (OPS)

Die Erstellung der Referenztablette erfolgte unter Verwendung der maschinenlesbaren Fassung des [Operationen- und Prozedurenschlüssels \(OPS\) \(Version 2022\) des Bundesinstituts für Arzneimittel und Medizinprodukte \(BfArM\)](#). Die Nutzungsbedingungen der Operationen- und Prozedurenschlüssels (OPS) des BfArM [hier](#) hinterlegt.

[Klassifikationen/ops.csv](#)

Variablen und Ausprägungen der Referenztablette:

Variable	Typ	Ausprägungen	Beschreibung
id	String	z. B. 1-202	ID, maximal 6-stellig
chapter	Integer	z. B. 1	OPS-Kapitel
group	String	z. B. 1-20 - 1-33	OPS-Gruppe, -Bereich
code3	String	z. B. 1-20	OPS-Kategorie/-Kode, 3-stellig
code	String	z. B. 1-202	OPS-Kategorie/-Kode, 4-stellig
code5	String	z. B. 1-202.-	OPS-Kategorie/-Kode, 5-stellig
code6	String	z. B. 1-202.--	OPS-Kategorie/-Kode, 6-stellig
name	String	z. B. Diagnostik zur Feststellung ...	Klassentitel der Maßnahme

Therapieprotokoll

Bei der verwendeten Referenztablette handelt es sich um eine *Vorschlagsliste* der [Plattform § 65c](#). Diese Vorschlagsliste stellt keine verbindliche Festlegung dar. Ein anerkannter Standard für die Kodierung von Systemtherapie-Protokollen ist uns nicht bekannt. Vorschläge für eine standardisierte Nomenklatur ([Rubinstein et al, 2020](#)), Referenzsysteme ([HemOnc.org](#), [National Cancer Institute Thesaurus \(NCIT\)](#)) und kommerzielle Produkte für die medizinische Dokumentation ([Onkopti®](#)) wurden von anderen entwickelt. Wir verweisen hier auf eine Auswahl dieser Arbeiten und Systeme.

Rubinstein, S. M., Yang, P. C., Cowan, A. J., & Warner, J. L. (2020). Standardizing Chemotherapy Regimen Nomenclature: A Proposal and Evaluation of the HemOnc and National Cancer Institute Thesaurus Regimen Content. *JCO clinical cancer informatics*, 4, 60–70. <https://doi.org/10.1200/CCI.19.00122>

[Onkopti® - die Datenbank digitalisierter onkologischer Therapieprotokolle](#)

Verwendete Substanzen

Die Erstellung der Referenztafelte erfolgte unter Verwendung des [Umsetzungsleitfadens Substanzen](#) in der Version 2021. Dieser Leitfaden wurde geringfügig, überarbeitet, in folgenden Schritten:

- der Eintrag [Larotrectinib](#) liegt als Duplikat vor, dies wurde korrigiert
- die Spalte [Code](#) wurde ergänzt, hier sind gültige ATC-Codes verknüpft, sofern anwendbar
- für die Zuordnung wurde auf das Kapitel [L](#) beschränkt (*Antineoplastic and immunomodulating agents*)

[Klassifikationen/substanz.csv](#)

Variablen und Ausprägungen der Referenztafelte:

Variable	Typ	Ausprägungen	Beschreibung
Therapieart	String	z. B. H0	Art der Therapie
Substanz	Sting	z. B. Abacavir	Bezeichnung des Arzneimittels
Code	String	z. B. J05AF06	ATC-Kode, Ebene 5

TNM: Regionäre Lymphknotenmetastasierung

Die Variable beschreibt den Status der regionären Lymphknotenmetastasierung (N-Kategorie der TNM).

[Klassifikationen/tnm_n.csv](#)

Zusatz (1mi), Mammakarzinom

Anwendung bei: Mikrometastase(n), > 0,2 mm und/oder mehr als 200 Tumorzellen, aber nicht größer als 0,2 cm

Stadium IB nach TNM8: T0, T1 N1mi M0

Quellen:

[Kapitel 11.3, Interdisziplinäre S3-Leitlinie für die Früherkennung, Diagnostik, Therapie und Nachsorge des Mammakarzinoms \(2021\);](#)

[TNM Classification of Malignant Tumours, 8th edition](#)

Zusatz (sn)

Ausprägung	Beschreibung
(p)NX(sn)	Schildwächterlymphknoten kann histologisch nicht beurteilt werden
(p)N0(sn)	Histologisch keine Lymphknotenmetastasen in Schildwächterlymphknoten
(p)N1(sn)	Befall des (der) Schildwächterlymphknoten

Quelle: [TNM Classification of Malignant Tumours, 8th edition](#)

Zusatz (i+), (mol+)

Ausprägung	Beschreibung
(p)N0	Histologisch keine Lymphknotenmetastasen, keine Untersuchung zum Nachweis isolierter Tumorzellen
(p)N0(i-)	Histologisch keine Lymphknotenmetastasen, kein morphologischer Nachweis von isolierten Tumorzellen
(p)N0(i+)	Histologisch keine Lymphknotenmetastasen, morphologischer Nachweis von isolierten Tumorzellen
(p)N0(mol-)	Histologisch keine Lymphknotenmetastasen, kein nichtmorphologischer Nachweis von isolierten Tumorzellen
(p)N0(mol+)	Histologisch keine Lymphknotenmetastasen, nicht-morphologischer Nachweis von isolierten Tumorzellen

Quelle: [TNM Classification of Malignant Tumours, 8th edition](#)

Beispieldaten

In diesem Repository soll der Bereitstellungsprozess für klinische Daten veranschaulicht werden.

Zum einen ist ein XML-Rohdatensatz hinterlegt für die Lieferung der Daten aus den klinischen Krebsregistern der Länder. Dieser entspricht den gemeinsam erarbeiteten Vorgaben des `oBDS-RKI` und wird im ZfKD zu einem deutschlandweiten Gesamtdatensatz verarbeitet. Der "rohe" Datensatz bestehend aus XML-Daten und bildet den Ausgangspunkt der weiteren Verarbeitung, wird aber vom ZfKD *nicht* ausgegeben.

Zum anderen wird hier simuliert, wie eine definierte Teilmenge des verarbeiteten Gesamtdatensatzes auf Antrag übermittelt wird. Zur Veranschaulichung dieser Datenbereitstellung dient der Ordner `Beispieldaten/csv`. Werden im Rahmen einer Antragsverfahren im ZfKD Daten übermittelt, entsprechen diese in Form und Aufbau exakt den hier abgelegten Beispieldateien. Die csv-Dateien in diesem Ordner sind aus dem XML-Rohdatensatz generiert.

Dabei ist zu beachten, dass aufgrund der vielfältigen Beziehungen der Bestandteile im [Datenschema](#) die Integration in eine einzelne Tabelle / Datei meist nicht zielführend ist. Es werden daher die angefragten Entitäten als einzelne Tabellen / Dateien exportiert. Die hier verwendeten Identifikatoren können in einem relationalen Modell wieder korrekt zusammengeführt werden (so sind etwa Einträge der Tumortabelle den jeweiligen Patienten zuordenbar). Hilfestellung bei den relationalen Beziehungen bietet das angehangene [ER-Modell](#), als strukturiertes Konzept zur Darstellung und Analyse von Daten in einer Datenbank.

Die in den Beispieldateien hinterlegten Daten sind künstlich erzeugt, folgen einfachen Verteilungen und berücksichtigen keine medizinischen Zusammenhänge. Die Identifikatoren sind zufällig erzeugt. Es besteht daher keinerlei Verbindung zu realen Daten.

Datei	Beschreibung	Download
Rohdatensatz	Ein einfacher Testdatensatz als <code>.xml</code> -Datei, bestehend aus Angaben zu 30 fiktiven Patienten.	
Applikationsart	Angaben zu Applikationsarten als <code>.csv</code> Testdatensatz mit Referenz auf <code>Bestrahlung</code> .	
Bestrahlung	Angaben zu Bestrahlungen als <code>.csv</code> Testdatensatz mit Referenz auf <code>ST</code> .	
FM	Angaben zu Fernmetastasen (Tumor) als <code>.csv</code> Testdatensatz mit Referenz auf <code>Tumor</code> .	
Folgeereignis FM	Angaben zu Fernmetastasen als <code>.csv</code> Testdatensatz mit Referenz auf <code>Folgeereignis</code> .	
Folgeereignis Weitere Klassifikationen	Angaben zu Weitere Klassifikationen als <code>.csv</code> Testdatensatz mit Referenz auf <code>Folgeereignis</code> .	
Folgeereignis	Angaben zu Folgeereignissen als <code>.csv</code> Testdatensatz mit Referenz auf <code>Tumor</code> .	
OP	Angaben zu Operationen als <code>.csv</code> Testdatensatz mit Referenz auf <code>Tumor</code> .	
	Angaben zu OP-Kodes als <code>.csv</code> Testdatensatz	

OPS	mit Referenz auf <code>OP</code> .	
Patient	Angaben zu Patienten als <code>.csv</code> Testdatensatz.	
Protokoll	Angaben zu Therapieprotokollen als <code>.csv</code> Testdatensatz mit Referenz auf <code>SYST</code> .	
ST	Angaben zu Strahlentherapien als <code>.csv</code> Testdatensatz mit Referenz auf <code>Tumor</code> .	
Substanz	Angaben zu Substanzen als <code>.csv</code> Testdatensatz mit Referenz auf <code>SYST</code> .	
SYST	Angaben zu systemischen Therapien als <code>.csv</code> Testdatensatz mit Referenz auf <code>Tumor</code> .	
Todesursache	Angaben zu Todesursachen als <code>.csv</code> Testdatensatz mit Referenz auf <code>Patient</code> .	
Tumor	Angaben zu Tumoren als <code>.csv</code> Testdatensatz mit Referenz auf <code>Patient</code> .	
Weitere Klassifikationen	Angaben zu Weitere Klassifikationen als <code>.csv</code> Testdatensatz mit Referenz auf <code>Tumor</code> .	

Metadaten

Zur Erhöhung der Auffindbarkeit sind die bereitgestellten Daten mit Metadaten beschrieben. Über GitHub Actions werden Metadaten an die entsprechenden Plattformen verteilt. Für jede Plattform existiert eine spezifische Metadatenfile, diese sind im Metadatenordner hinterlegt:

[Metadaten/](#)

Versionierung und DOI-Vergabe erfolgt über [Zenodo.org](https://zenodo.org). Die für den Import in Zenodo bereitgestellten Metadaten sind in der `zenodo.json` hinterlegt. Die Dokumentation der einzelnen Metadatenvariablen ist unter <https://developers.zenodo.org/representation> nachlesbar.

[Metadaten/zenodo.json](#)

Hinweise zur Nachnutzung der Daten

Offene Forschungsdaten des RKI werden auf [Zenodo.org](https://zenodo.org), [GitHub.com](https://github.com), [OpenCoDE](https://opencode.rki.de) und [Edoc.rki.de](https://edoc.rki.de) bereitgestellt:

- <https://zenodo.org/communities/robertkochinstitut>
- <https://github.com/robert-koch-institut>

- <https://gitlab.opencode.de/robert-koch-institut>
- <https://edoc.rki.de/>

Lizenz

Der Datensatz "Bundesweiter klinischer Krebsregisterdatensatz - Datenschema und Klassifikationen" ist lizenziert unter der [Creative Commons Namensnennung 4.0 International Public License | CC-BY](#) .

Die im Datensatz bereitgestellten Daten sind, unter Bedingung der Namensnennung des Robert Koch-Instituts als Quelle, frei verfügbar. Das bedeutet, jede Person hat das Recht die Daten zu verarbeiten und zu verändern, Derivate des Datensatzes zu erstellen und sie für kommerzielle und nicht kommerzielle Zwecke zu nutzen. Weitere Informationen zur Lizenz finden sich in der [LICENSE](#) bzw. [LIZENZ](#) Datei des Datensatzes.