

CARLOS

Cancer Registry
Lower-Saxony



Das Datenschema des Epidemiologischen Krebsregisters Niedersachsen für die Erprobungsphase

- Dritte, überarbeitete Auflage -

**Dipl.-Math. A. Scharnofske
Dipl.-Inform. F. Wietek**

Oldenburg, 16.2.1998

Inhaltsverzeichnis

Inhaltsverzeichnis.....	i
Abbildungsverzeichnis.....	iii
Vorwort	v
1. Die Vertrauensstelle	1
1.1 Das ER-Schema	1
1.2 Relationen	5
1.2.1 Meldungen	5
1.2.2 Weitere meldungsbezogene Angaben.....	14
1.2.3 Klartexte	17
1.2.4 Personenidentifizierende Daten, Kontrollnummern und Chifftrate.....	19
1.2.5 Kodierungstabellen für Attribute, Dokumentationskräfte und Melder.....	21
1.2.6 Hilfsrelationen	28
1.2.7 Relationen für Klartextabgleich und Verwaltung.....	29
2. Die Registerstelle	33
2.1 Das ER-Schema	33
2.2 Relationen	37
2.2.1 Patienten	39
2.2.2 Tumoren.....	40
2.2.3 Meldungen	45
2.2.4 Weitere meldungsbezogene Daten.....	50
2.2.5 Texte und Chifftrate.....	52
2.2.6 Kodierungstabellen für Attribute, Dokumentationskräfte und Melder.....	53
2.2.7 Hilfsrelationen	54
2.2.8 Relationen für Datensatzabgleich und Verwaltung	55
2.3 Indizes.....	58
3. Unterschiede zum Standard der Arbeitsgemeinschaft bevölkerungsbezogener Krebsregister	61
3.1 Grundsätzliche Überlegungen.....	61
3.2 Personenbezogene Angaben	61
3.3 Tumorbezogene Angaben	62
3.4 Angaben zur Therapie.....	63
3.5 Angaben zum Sterbefall.....	63
4. Geodaten in der Registerstelle	65
4.1 Konzeptionelles ER-Schema	65

4.2	Relationen	69
4.2.1	Verwendete Domänen	69
4.2.2	Tabellen	70
4.2.3	Untersuchungsgebiete.....	73
4.2.4	Regionalinformationen	78
4.2.5	Standardbevölkerungen	80
4.3	Indizes	83
Anhang A: Kodierungen.....		85
A.1	Die Internationale Klassifikation der Krankheiten (ICD).....	85
A.1.1	9. Revision der ICD	85
A.1.2	10. Revision der ICD	86
A.2	Die ICD für die Onkologie (ICD-O).....	87
A.2.1	1. Auflage der ICD-O	88
A.2.2	2. Auflage der ICD-O	89
A.2.3	Deutsche Auflage der ICD-O (ICD-O-DA).....	89
A.3	TNM.....	90
A.4	Berufskodierung.....	92
A.5	Staatenkodierung.....	93
A.6	Gemeindegkennziffern und regionale Beobachtungseinheiten.....	93
A.7	Gauß-Krüger-Koordinaten	94
Anhang B: Datenbank-Installation		97
B.1	Verzeichnisstruktur	97
B.2	Schemainstallation	98
B.2.1	Wertebereiche der Attribute	99
B.2.2	Schemainstallation der Kodierungstabellen	99
B.2.3	Schemainstallation der Geo-Daten	100
B.2.4	Schemainstallation der Fall- und Meldungsdaten	100
B.2.5	Schemainstallation für die Bevölkerungsdaten	101
B.3	Laden der Daten	101
B.3.1	Vollständiges Laden der Daten.....	102
B.3.2	Änderungen an den geladenen Daten	106
B.4	Sonstige Verzeichnisse.....	111
B.4.1	Füllen der Relation Erweiterte_Tumordaten	111
B.4.2	Hilfsskripte	112
Literaturverzeichnis.....		113

Abbildungsverzeichnis

Abb. 1.1: ER-Schema der Vertrauensstelle: Meldungen.....	1
Abb. 1.2: ER-Schema der Vertrauensstelle: Unklarheiten, Anfragen und Vitalstatusmeldung.....	3
Abb. 2.1: Konzeptionelles ER-Schema der Registerstelle	34
Abb. 2.2: Implementiertes ER-Schema der Registerstelle: Meldungen.....	35
Abb. 2.3: Implementiertes ER-Schema der Registerstelle: Abgleich, Unklarheiten, Anfragen und Vitalstatusmeldung.....	36
Abb. 4.1: Konzeptionelles ER-Schema für Geometrieobjekte.....	66
Abb. 4.2: Konzeptionelles ER-Schema für Untersuchungsgebiete	67
Abb. 4.3: Beziehungen zwischen Gebieten und Geo-Objekten	68
Abb. 4.4: Standardbevölkerungen mit Fallzahlen	81

Vorwort

Dieser Bericht stellt das epidemiologische Datenschema des Epidemiologischen Krebsregisters Niedersachsen zum fortgeschrittenen Stand der Erprobungsphase vor (vgl.[AFH+97, AFH+96, AFK+95]) und bildet somit eine Überarbeitung von [Wie96, WS97]. Da sich viele Meldewege weiterhin noch im Aufbau befinden und noch Struktur und Umfang von Meldebögen (insb. des Epidemiologischen Krebsregisters selbst) diskutiert werden, kann dieses Schema nur vorläufigen Charakter haben. Dies äußert sich auch in den Wertebereichen einiger Attribute, die über das Angebot möglichst vieler verschiedener Ausprägungen bzw. mit der Möglichkeit, beliebige Freitexte abzulegen, das endgültige Format noch teilweise offen lassen, um zunächst möglichst alle gemeldeten Daten speichern zu können und den Aufwand für Umstrukturierungen des Schemas gering zu halten. Weiterhin steht auch noch die Verabschiedung eines Landeskrebsregistergesetzes an, aus dem sich eventuell noch Modifikationen am Datenschema ergeben könnten.¹

Insgesamt wurde Wert darauf gelegt, alle Meldungen möglichst vollständig und originalgetreu in das Datenschema übernehmen zu können. Eine geeignete Weiterverarbeitung der aus den Meldungen gewonnenen Angaben im Rahmen eines Abgleichs zwischen verschiedenen Meldungen zu ein und demselben Patienten bzw. Tumor erfolgt noch nicht in allen Fällen. Diesbezügliche Überlegungen auf der Basis der Erfahrungen mit erfaßten Datensätzen und auch im Hinblick auf den noch genauer festzulegenden Katalog an Auswertungsmöglichkeiten in der Registerstelle (etwa im Hinblick auf den Jahresbericht des Registers) müssen noch erfolgen und können ggf. noch zu einer Überarbeitung und Erweiterung des Datenschemas führen.

In diesem Bericht wurde auf den Überblick über den betrachteten epidemiologischen Datensatz verzichtet (siehe [Wie96]). Neu gegenüber [WS97] ist die Möglichkeit, voraggregierte Fallzahlen zur Berechnung externer Standards abzulegen. Neben einzelnen Ergänzungen und Modifikationen wurde die Beschreibung der Datenbank-Installation in der Registerstelle überarbeitet.

Im einzelnen stellen Kapitel 1 und 2 das Datenschema der Vertrauensstelle sowie den epidemiologischen Teil des Datenschemas der Registerstelle im Detail vor. Die Schemata sind aufeinander abgestimmt und stützen sich in der Grundlage auf die Anforderungen des Bundeskrebsregistergesetzes [KRG94] sowie in der Umsetzung im wesentlichen auf die Arbeiten der GMDS-Projektgruppe „Krebsregister“ bzw. der Arbeitsgemeinschaft bevölkerungsbezogener Krebsregister in Deutschland. In einzelnen Punkten weichen sie jedoch leicht von dem dort erarbeiteten Vorschlag „Dokumentationsstandards“ [Sch96, SAW95] ab – siehe hierzu Kapitel 3. Kapitel 4 geht auf die Modellierung der Geodaten ein. Im Anhang werden schließlich spezielle im Schema verwendete Domänen (ICD- und ICD-O-Kodierung, TNM-Klassifikation, Berufs-, Staaten- und Regionenkodierung sowie Gauß-Krüger-Koordinaten) definiert und die Installation des Datenschemas der Registerstelle beschrieben.

Für die nächste Version des Datenschemas (erscheint voraussichtlich Sommer 1998) werden größere Änderungen im Bereich der Auswertungsrelationen der Registerstelle anstehen. Die Verwaltung der multidimensionalen Daten (im wesentlichen die Fall- und Einwohnerdaten) wird zukünftig für einen schnelleren Zugriff optimiert werden. Weiterhin sollen Zwischenergebnisse zur späteren Wiederverwendung gesichert werden. Die Geo-Daten werden auf das GIS-System

¹ So wird etwa das Führen des Ausfülldatums einer Meldung angedacht.

des InterGIS-Projekts [Fri97] verlegt und aus der Datenbank entfernt. Die Hierarchie- und Regionalinformationen werden zwar weiterhin in der Datenbank bleiben, müssen jedoch reorganisiert werden.

1. Die Vertrauensstelle

1.1 Das ER-Schema

Im Zentrum des Datenschemas der Vertrauensstelle steht die Meldung. Die Relation Meldungen enthält lediglich verschiedene zur Verwaltung der gespeicherten Meldungen benötigte Angaben, während in den Relationen, die die Meldung weiter spezialisieren, alle durch den jeweiligen Meldebogen bzw. Melder gelieferten epidemiologischen Daten vor der Weiterleitung an die Registerstelle möglichst vollständig und originalgetreu gespeichert werden. Nach Abschluß der Bearbeitung einer Meldung in der Registerstelle, spätestens aber 3 Monate nach Übermittlung an die Registerstelle wird die Meldung in der Vertrauensstelle gelöscht.

Jede Meldung ist einem meldenden Arzt oder einem Sammelmelder (einer „im Auftrag“ meldenden Einrichtung), etwa einer Nachsorgeleitstelle, einem Klinikregister, Tumorzentrum oder Gesundheitsamt, zugeordnet. Auch für Meldungen von Sammelmeldern sollte der ursprünglich meldende Arzt definiert sein. Zusätzlich enthalten Sterbemeldungen die separate Angabe des behandelnden oder Hausarztes. Der Datenbestand über Melder wird im Grunde genommen unabhängig von konkreten Meldungen geführt und sowohl durch Absprache mit potentiellen Meldern als auch aufgrund von Angaben auf Meldungen ergänzt.

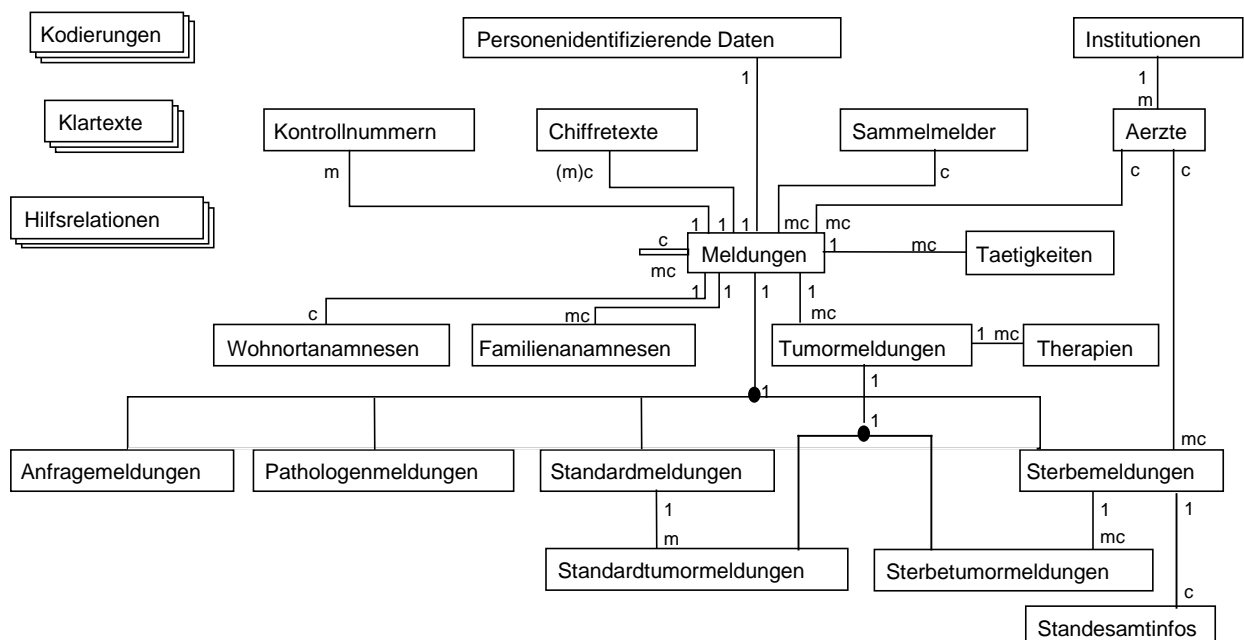


Abb. 1.1: ER-Schema der Vertrauensstelle: Meldungen

Weiterhin existieren zu einer Meldung ein Satz personenidentifizierender Angaben, der nicht an die Registerstelle weitergereicht wird, eine Menge von hieraus abgeleiteten Kontrollnummern, die in der Registerstelle den personenbezogenen Abgleich mit anderen Meldungen ermöglichen, sowie i.a. auch der jeweils zugehörige chiffrierte personenidentifizierende Datensatz (Chiffretext)², der auch zur dauerhaften Speicherung in der Registerstelle vorgesehen ist. Eine Ausnahme bilden hier die faktisch anonymisiert weiterzureichenden Pathologenmeldungen.³

In einer Meldung können verschiedene, (im Rahmen einer Untersuchung) gleichzeitig diagnostizierte Tumoren (etwa auch multiple Tumoren) in Form von Tumormeldungen beschrieben werden – Tumoren mit unterschiedlichen Diagnosedaten werden also stets auf mehrere Meldungen (mit jeweils einer Tumormeldung) verteilt. Aufgrund von aussagekräftigen Klartextangaben zu früheren Tumoren in einer Meldung kann die Vertrauensstelle selbst neue Meldungen über diese Tumoren generieren. Diese Meldungen tragen einen Verweis auf die „Ursprungsmeldung“, werden aber ansonsten so behandelt, als wären sie vom jeweiligen Melder separat eingesandt worden. Eine Meldung kann generell Angaben zu verschiedenen Tätigkeiten des Patienten bis zum Zeitpunkt der Meldung, früheren Wohnorten, Krebsfällen in der Familie sowie den Komponenten der Primärtherapie jedes gemeldeten Tumors machen.

Es werden

- Standardmeldungen zu Neuerkrankungen mit erfolgter Einwilligung oder über eine Ausnahmeregelung für Patienten, die verstorben oder nicht nur vorübergehend nicht einwilligungsfähig sind bzw. denen die Aufklärung über ihr Krebsleiden nicht zuzumuten ist,
- Pathologenmeldungen, die einen etwas reduzierten, auf die Anonymisierung der Angaben bedachten Datensatz liefern, sowie
- Sterbemeldungen über Todesfälle von Krebspatienten

unterschieden.⁴ Gemäß dem Meldeweg der „peripheren Teilanonymisierung mit zentraler Verschlüsselung“ (vgl. [AFK+95]) werden bei Pathologenmeldungen die epidemiologischen, also tumorbezogenen Daten direkt vom Melder an die Registerstelle übertragen und somit nicht in der Vertrauensstelle gespeichert. Sterbemeldungen bestehen derzeit nur vor allem Totenscheinen aus dem Gesundheitsamt, aber auch klinischen Abschlußmeldungen. Zu jedem Totenschein existiert eine Sterbebuchnummer eines Standesamtes, die in den Standesamtinfos vermerkt ist. Zu Totenscheinen ohne Angaben von Krebs sollen jeweils zunächst nur die für den Abgleich in der Registerstelle nötigen Informationen extrahiert und weitergeleitet werden. Gibt es zum betroffenen Patienten bereits Meldungen in der Registerstelle, fragt diese (in Form einer Unklarheit – s.u.) bei der Vertrauensstelle die vollständigen Informationen nach. Andernfalls wird der Totenschein wieder aus dem Registerbestand gelöscht. Damit bei diesem Vorgehen möglichst keine relevanten Todesbescheinigungen fälschlicherweise gelöscht werden, sollten Totenscheine jeweils erst *nach* allen Inzidenzmeldungen des betreffenden Zeitraums mit der Registerdatenbank abgeglichen werden.

² In der Implementierung ist der Chiffretext auf mehreren Zeilen aufgeteilt – dies wird durch das (m) im ER-Diagramm repräsentiert.

³ Von Pathologenmeldungen wird im folgenden vereinfachend immer gesprochen, wenn von faktisch anonymisierten Meldungen gemäß dem derzeitigen Entwurf zum Niedersächsischen Krebsregistergesetz die Rede ist. Hierunter fallen insbesondere auch Meldungen aus Berufskrankheitenverfahren.

⁴ Details der Datensätze der einzelnen Melder sowie das genaue Vorgehen bei der Auswahl und Konvertierung von Attributen aus den gemeldeten Daten in das Datenschema des Krebsregisters werden in einem gesonderten Bericht beschrieben werden.

Zur temporären Repräsentation von Auskunftsanforderungen und Widerrufen von Einwilligungen durch Patienten dienen Anfragemeldungen, die nur aus personenidentifizierenden Daten bzw. Kontrollnummern und einigen wenigen patientenbezogenen Attributen zum Abgleich mit den Registermeldungen bestehen.

Weiterhin werden im Datenschema gemeldete Klartexte verschiedener Typen (Diagnose, Berufe, Wohnorte, allgemeine Kommentare usw.) in separaten Tabellen abgelegt, die den Meldungen jeweils über Fremdschlüssel zugeordnet sind. Zu verschiedenen Attributen werden Kodierungstabellen vorgehalten, die zu den jeweiligen numerischen Ausprägungen Volltexte sowie mnemotechnische Codes zur Verwendung in der Benutzungsoberfläche des Systems bzw. zur Unterstützung von Transfers von und zu anderen Systemen definieren.

Abbildung 1.1 zeigt in einem Entity-Relationship-Diagramm (ER-Diagramm) die zentralen Objektklassen (Entitäten) des Datenschemas rund um die Meldung sowie die Beziehungen zwischen diesen. Zu den Beziehungen ist jeweils die Anzahl der beteiligten Entitäten angegeben: „1“ steht für „genau ein“, „c“ für „höchstens ein“, „m“ für „mindestens ein“ und „mc“ für „kein, ein oder mehrere“.

In Abbildung 1.2 werden die Datenbanktabellen vorgestellt, die zu Verwaltungszwecken (im Zuge von Rückfragen aus der Registerstelle, Auskunftsanforderungen und Einwilligungswiderrufen durch Patienten sowie zur Behandlung von Rückmeldungen des Vitalstatus von Patienten an die Melder) – vor allem temporär zur Kommunikation mit der Registerstelle – dienen. Hierbei sind Entitäten und Beziehungen, die bereits in Abbildung 1.1 dargestellt wurden, hervorgehoben.

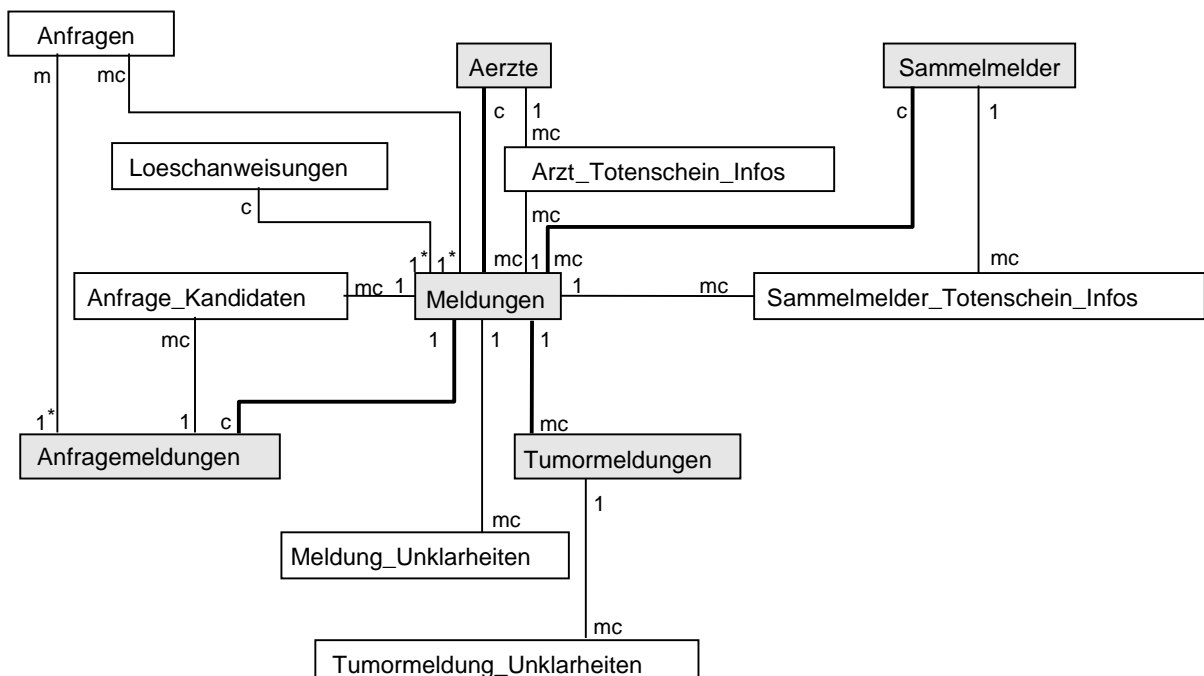


Abb. 1.2: ER-Schema der Vertrauensstelle: Unklarheiten, Anfragen und Vitalstatusmeldung

Falls bereits bei der Meldungsbearbeitung in der Vertrauensstelle oder als Ergebnis der Abgleichsaufbereitung in der Registerstelle Unklarheiten auftreten, wird dies durch Einträge in den beiden Relationen *Meldung_* und *Tumormeldung_Unklarheiten* vermerkt. Bei einer Unklarheit in der Registerstelle wird die Meldung (mit allen zugehörigen Tumormeldungen und ggf. auch den aus ihr durch die Vertrauensstelle abgeleiteten Meldungen zu früheren Tumoren) dort gelöscht und muß nach Klärung aller Fragen komplett neu gemeldet werden. Auch die erfolgreiche Bearbeitung einer Meldung in der Registerstelle (und damit der Aufruf zur Löschung der Meldung in der Vertrauensstelle) wird über diese Relationen gemeldet, wobei für DCO-Fälle zusätzlich zum Follow-Back aufgerufen wird. Weiterhin werden auch die Ergebnisse von Nachfragen temporär hier dokumentiert. Zu Pathologenmeldungen ist keine Rückfrage aus der Registerstelle möglich.

Zu Anfrage-, also Auskunfts- und Widerspruchsmeldungen werden Anfragekandidaten gemäß dem automatischen Abgleich durch Automatch aus dem Meldungsbestand der Registerstelle für den Klartext-Abgleich in der Vertrauensstelle gekennzeichnet. Pathologenmeldungen sind hier, da keine Einwilligung vorliegt, aus Gründen des Datenschutzes ausgenommen. Je nach Fragestellung kann bzw. muß (bei fehlender Genehmigung) für Anfragen aus der Forschung auf den Klartextabgleich verzichtet werden. Um auch die Berücksichtigung von noch in der Bearbeitung befindlichen betroffenen Meldungen zu gewährleisten, wartet die Vertrauensstelle mit der Bearbeitung von Patientenauskünften und Widerrufen so lange, bis alle Meldungen mit einem früheren Eingangsdatum an die Registerstelle übersandt wurden. Für Anfragen aus der Forschung kann diese Frist ggf. entfallen. Die Registerstelle ermittelt die Kandidatenliste dann aus allen (auch den noch nicht abgeglichenen oder nicht vollständig nachbearbeiteten) Meldungen und transferiert dann lediglich die bereits in der Vertrauensstelle gelöschten Meldungen komplett zurück.

Nach dem Klartextabgleich von Kandidaten zu einer Widerspruchsmeldung werden in Löschanweisungen die in der Registerstelle zu löschenden Meldungen aufgelistet, wobei Homonymfehler berücksichtigt, d.h. ehemals fälschlicherweise dem betroffenen Patienten zugeordnete Meldungen nicht aus dem Registerbestand gelöscht werden.⁵ Nach dem Löschen muß ggf. ein neues Best-of über verbleibende Meldungen in der Registerstelle durchgeführt werden. Über erfolgte Auskünfte und Löschungen inklusive der betroffenen Registermeldungen wird dauerhaft Buch geführt.⁶

Zu verstorbenen Patienten können weiterhin mit Bezug auf eine Sterbemeldung nach Abgleich und Meldungsbearbeitung in der Registerstelle im Rahmen eines genehmigten Forschungsvorhabens Melder spezifiziert werden, die über den Tod zu informieren sind. (Hierzu dienen die Relationen *Arzt_* bzw. *Sammelmelder_Totenschein_Infos*.) Dies betrifft nur Melder, die ehemals zum jetzt verstorbenen Patienten Meldungen übermittelt haben. Es wird kein Klartextabgleich durchgeführt – evtl. treten also Homonymfehler und somit nicht erforderliche Benachrichtigungen auf.

Das genaue Vorgehen bei den angeführten Verwaltungsaufgaben sowie die exakte Nutzung der angebotenen Relationen wird in dem Bericht [HDW97] beschrieben.

⁵ Eine Korrektur derartiger Fehler im Datenbestand der Registerstelle im Rahmen der Bearbeitung von Auskunftsanforderungen ist aus Gründen des Datenschutzes (Zweckbindung der Dechiffrierung) nicht vorgesehen.

⁶ Da die Protokollierung von Anfragen sowie temporär (bis zur Absendung an die Registerstelle) auch von Löschanweisungen die jeweils referenzierten Meldungen und Anfragemeldungen überdauert, ist hier im Schema keine referentielle Integrität gewährleistet. (Dies ist in Abb. 1.2 durch ein „*“ gekennzeichnet.)

1.2 Relationen

Im folgenden werden die einzelnen Relationen des Datenschemas mit ihren jeweiligen Attributen genauer vorgestellt.⁷ Zu jedem Attribut sind in tabellarischer Form jeweils der ORACLE-Datentyp, mögliche Ausprägungen und evtl. kurze Kommentare angegeben. Am Beginn einer Tabelle stehen stets (unterstrichen) die Primärschlüssel-Attribute, gefolgt von eventuellen Fremdschlüsseln und dann allen weiteren Angaben. Die fortlaufende Numerierung bei Schlüsselattributen beginnt jeweils ab 1.

Soweit nicht explizit erwähnt, sind NULL-Werte im Schema nicht vorgesehen, sondern werden durch eine entsprechende Ausprägung repräsentiert (z.B. in der ICD-Kodierung 239.9 bzw. D48.9 oder oft als „9 – fehlende Angabe / unbekannt“).

Nähere Anmerkungen zu den einzelnen Attributen beschränken sich hier im wesentlichen auf allgemeine an die Melder gerichtete Instruktionen bzw. die zu meldenden Daten betreffende Hinweise. Eine entsprechende detailliertere und melderspezifische Beschreibung der Datentransfers wird in einem gesonderten Bericht erscheinen.

1.2.1 Meldungen

Relation Meldungen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Fortlaufende Nummer.
meldung_ref	number(8)	Für Meldungen, die in der Vertrauensstelle gemäß Klartextangaben zu früheren Tumoren aus einer anderen Meldung definiert wurden, eine Referenz auf diese – sonst NULL.
arzt_ref	number(5)	Fremdschlüssel aus Ärzte – ggf. NULL.
sammelmelder_ref	number(3)	Fremdschlüssel aus Sammelmelder – ggf. NULL.
vst_bearbeiter_ref	number(2)	Fremdschlüssel aus VST_Bearbeiter. Der Erfasser des Meldebogens bzw. Bearbeiter der Meldung.
meldungsstatus	number(1)	- 1 (Neumeldung) - 2 (in Bearbeitung) - 3 (in Nachfrage) - 5 (Registermeldung) Folgende Ausprägung (vgl. Abschnitt 2.2.3) wird in der Vertrauensstelle nicht benötigt: - 4 (vorläufige Registermeldung)

⁷ Nicht weiter eingegangen wird auf durchgängig in ORACLE definierte Primär- und Fremdschlüssel- sowie CHECK-Constraints, die die Wertebereiche von Attributen vom Aufzählungstyp überprüfen, soweit diese zum Zeitpunkt der Schemadefinition bereits klar sind und nicht erst durch später geladene Tabellen definiert werden. (Dies ist für alle in Abschnitt 1.2.5 separat aufgeführten Attribute, beginnend bei den ICD-Codes, der Fall.)

meldemodus	number(2)	- 1 (mit Information) - 2 (Ausnahmeregelung) - 3 (Pathologe) - 4 (Berufskrankheitenverfahren) - 5 (Todesbescheinigung) - 6 (Autopsiebefund) - 7 (klinischer Abschluß) - 10 (Widerspruchsmeldung) - 11 (Auskunftsanforderung Patient) - 12 (Auskunftsanforderung Forschung)
externe_referenz	char(15)	Externe Referenznummer, die nur für den Melder eine Semantik trägt (Identifikation des Patienten oder sogar der Meldung) – ggf. NULL, falls nicht benötigt bzw. erlaubt.
eingangsdatum	date	Eingang in die Vertrauensstelle (Meldebogen oder Datenträger).
erfassungsdatum	date	Aufnahme in die Datenbank der Vertrauensstelle.
registrierungsdatum	date	Übersendung an die Registerstelle – vorher NULL.

Alle Meldungen werden fortlaufend durchnummeriert.

Falls in einer Meldung detaillierte Klartexte zu früheren Tumoren angegeben sind, die – nach Einschätzung der bearbeitenden Dokumentationskraft – eine fundierte und aussagekräftige Aussage erlauben, so können hieraus in der Vertrauensstelle neue Meldungen zu diesen Tumoren definiert werden. Die Attributwerte dieser Meldungen werden in der Relation Meldungen – bis auf die externe Referenz – aus der „Ursprungsmeldung“ übernommen. Weiterhin werden auch personenidentifizierende Daten, Kontrollnummern und Chiffretexte kopiert, so daß die neuen Meldungen im wesentlichen so behandelt werden können, als wären sie separat vom Melder eingekommen. Die Ursprungsmeldung darf jedoch zur Unterstützung der Bearbeitung von Unklarheiten nicht vor einer daraus abgeleiteten Meldung als „abschließend bearbeitet“ aus dem Datenbestand der Vertrauensstelle gelöscht und auch nicht ohne ihre abgeleiteten Meldungen von der Registerstelle nachgefragt werden.

Im allgemeinen muß zumindest eine der beiden Referenzen auf *Arzt* und *Sammelmelder* definiert sein. Eine Ausnahme bilden Pathologenmeldungen, da hier datenschutzrechtliche Bedenken gegen eine dauerhafte Speicherung des Melders bestehen. Idealerweise sollte der meldende Arzt immer bekannt sein (bei Meldungen über Sammelmelder ist dies der Arzt, der an die betreffende Institution gemeldet bzw. – im Fall von mehreren Meldern – derjenige der die Einwilligung des Patienten eingeholt hat), während der Sammelmelder natürlich für Meldungen von Einzelmeldern undefiniert bleibt.

Über den *Meldungsstatus* werden zum Transfer in die Registerstelle bereitstehende Neumeldungen, noch zu bearbeitende Neumeldungen, in der Nachfrage befindliche Meldungen sowie für Auskunfts- und Löschungszwecke rücktransferierte Registermeldungen unterschieden.

Der *Meldemodus* lehnt sich an das im derzeitigen Entwurf zum Niedersächsischen Krebsregistergesetz festgelegte Meldemodell an. 1 oder 2 steht für Standardmeldungen; 5, 6 oder 7 für Sterbemeldungen sowie 3 oder 4 für (evtl. um eine Berufsanamnese erweiterte) Pathologenmeldungen. Hierbei sind 4 und 6 zunächst noch nicht relevant; die Meldemodi 10 bis 12 dienen für Anfragemeldungen.

Das *Registrierungsdatum* kennzeichnet den Beginn der 3-Monatsfrist zur Speicherung von Meldungen in der Vertrauensstelle.

Die Protokollierung des *Bearbeiters* einer Meldung hilft, Änderungen in der Meldung nachzuvollziehen.. Der Bearbeiter repräsentiert – im Gegensatz zum Erfassungsdatum – die letzte Beschäftigung mit der Meldung (auch im Fall von Nachfragen aus der Registerstelle). Ob eine längere oder genauere Historie sinnvoll und datenschutzrechtlich akzeptabel ist, muß noch diskutiert werden. Automatisiert aus CAMEL übernommenen Datenträgermeldungen wird zunächst „CAMEL“ als Bearbeiter zugeordnet.

Eine *externe Referenz* darf bei Standardmeldungen, obwohl es von einigen Meldern gewünscht wird, nach dem aktuellen Entwurf des Niedersächsischen Krebsregistergesetzes nicht gespeichert werden. Im Fall der Pathologenmeldungen erfolgt jedoch über diese Referenz in der Registerstelle die Zusammenführung von epidemiologischen und personenidentifizierenden Daten, daher ist sie hier unverzichtbar und darf in der Registerstelle auch zu diesem Zweck temporär gespeichert werden.

ist ebenso wie das Führen *externer Referenzen* datenschutzrechtlich noch zu diskutieren

Relation Tumormeldungen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Fremdschlüssel aus Meldungen.
<u>tumormeldung_id</u>	number(1)	Fortlaufende Numerierung der Tumoren einer Meldung.
multipler_tumor_id	number(1)	Fortlaufende Numerierung der „unabhängigen“ Tumoren einer Meldung.
validitaet	number(1)	Validität der Angaben zum Tumor: - 1 (valide / unauffällig / noch unbearbeitet) - 2 (auffällig, aber in Nachfrage bestätigt) - 3 (ungewöhnlich und (trotz Nachfrage) ungeklärt) - 4 (invalide / „unmöglich“ und ungeklärt)

Alle Tumoren einer Meldung werden fortlaufend durchnummeriert.

Unter den verschiedenen Tumoren einer Meldung können ggf. Tumoren als zur gleichen Erkrankung gehörig gekennzeichnet werden. Dies wird dadurch repräsentiert, daß diese unter *Multipler-Tumor-ID* den gleichen Schlüssel erhalten. Ansonsten trägt dieses Attribut keine Semantik, muß also insbesondere auch nicht mit *Tumormeldung-ID* konsistent sein.

Eine zusätzliche Angabe zur *Validität* der Tumordaten, also zur Wahrscheinlichkeit der Attributkombination des Datensatzes, kann als Hilfestellung beim Best-of in der Registerstelle sowie zum Ausschluß unsicherer Fälle in Registerauswertungen dienen. Die Einstufung der Validität basiert auf Notwendigkeit und Erfolg eventueller Nachfragen bei in der manuellen Bearbeitung oder aufgrund automatischer Plausibilitätskontrollen als zumindest unwahrscheinlich eingestuft Fällen.⁸ Es besteht noch ein gewisser Diskussionsbedarf über die Art der Nutzung – und daraus resultierend die benötigten Ausprägungen – eines derartigen Attributs. Evtl. könnte auch die Vollständigkeit der jeweiligen Meldung miteinbezogen werden.

⁸ Inwiefern es überhaupt „unmögliche“ Fälle gibt, ob also eine Unterscheidung zwischen ungewöhnlich und invalide sinnvoll ist, ist noch zu diskutieren.

Relation Anfragemeldungen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
geschlecht	number(1)	- 1 (männlich) - 2 (weiblich) - 9 (fehlende Angabe / Sonstige)
geburtsjahr	number(4)	Vierstelliges Jahr – 0 für unbekannt.
geburtsmonat	number(2)	Zweistelliger Monat (1-12, 0 für unbekannt).
wohnort	number(10)	Wohnort zum Zeitpunkt der Anfrage: Gemeindekennziffer (vgl. Anhang) bzw. entsprechende Kodierung auf höherer Aggregationsebene, falls nicht genauer möglich – 0 für unbekannt. ⁹ Zukünftig Berücksichtigung regionaler Beobachtungseinheiten (vgl. Anhang), also z.B. Kodierung v. Ortsteilen.

Für Anfragemeldungen werden neben den Kontrollnummern noch die oben angeführten Attribute benötigt, um den Abgleich in der Registerstelle durchführen zu können.

Relation Standardmeldungen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
geschlecht	number(1)	<s. Anfragemeldungen>
geburtsjahr	number(4)	<s. Anfragemeldungen>
geburtsmonat	number(2)	<s. Anfragemeldungen>
verstorben	number(1)	- 1 (Patient verstorben) - 2 (Patient nicht verstorben) - 9 (fehlende Angabe / unbekannt)
staatsangehoerigkeit	number(3)	Aktuelle (letzte) Staatsangehörigkeit sowie für Deutsche nach Möglichkeit Angabe des Bundeslandes des aktuellen (letzten) Wohnortes. Kodierung nach der Klassifikation des Statistischen Bundesamtes – 999 für unbekannt.
staat_version	number(4)	Jeweils verwendete Version der Länderkodierungen – 0 für unbekannt.
mehrling	number(1)	- 0 (kein Mehrling) - 1 (eineiiger Mehrling) - 2 (zweieiiger Mehrling) - 3 (Mehrling – unbekannt, ob ein- oder zweieiig) - 9 (fehlende Angabe / unbekannt)
wohnort	number(10)	<s. Anfragemeldungen>, Wohnort zum Zeitpunkt der Meldung.
geocode_x	number(7)	Gauß-Krüger-Koordinaten (vgl. Anhang; angestrebt sind 100 x 100 Meter, also zwei schließende Nullen) – 0 für unbekannt. Die x- (Rechts-) Koordinate beginnt mit dem Bezugsmeridian – zur Effizienzsteigerung soll evtl. alles auf den dritten Meridian (9° ö.L.) umgerechnet und in dieser Form abgelegt werden.
geocode_y	number(7)	<s. geocode_x>

⁹ Wenn zumindest klar ist, daß der betreffende Patient aus dem Untersuchungsgebiet (also etwa Niedersachsen) stammt, so sollte dies auch so anstelle von 0 kodiert werden.

raucher_status	number(1)	Raucher-Status zum Zeitpunkt der Meldung: - 1 (Nichtraucher) - 2 (Exraucher) - 3 (Raucher) - 9 (fehlende Angabe / unbekannt)
ex_raucher_jahr	number(4)	Das Jahr an dem der Raucher (s.o.) mit dem Rauchen aufgehört hat – 0 für unbekannt sowie für Raucher und Nichtraucher.
patienten_bemerkung	number(1)	Hat der Patient zu seiner Krebserkrankung weitere Anmerkungen gemacht: - 1 (ja) - 2 (nein) - 9 (nicht befragt / fehlende Angabe / unbekannt)

In der Relation Wohnort_Texte findet sich ggf. ein Klartext zum Wohnort. Evtl. wird später (in der Dauerphase) auch der Text der Patientenbemerkungen separat gespeichert. Spezifikationen früherer Tumoren liegen ggf. in der Relation Fruehere_Tumoren_Texte.

Es gibt Melder, die nur jeweils den aktuellen Daten (wie z.B. der Wohnort) und die Erkrankungshistorie des Patienten kennen und diese immer vollständig melden. Diese Meldungen ließen sich bisher nicht als eine Standardmeldung speichern, da das Diagnosedatum in der Standardmeldung mit der Erkrankungshistorie kollidierte, die mehr als ein Diagnosedatum umfassen konnte. Zu dieser Historie sind die Daten (wie z.B. der Wohnort) eigentlich nur für den Zeitpunkt der Meldung gültig. Diese Daten können somit meist für ältere Erkrankungen nicht korrekt sein.

Wenn aber beispielsweise zu den einzelnen Diagnosezeitpunkten der Wohnort bekannt ist, dann kann diese Information nicht in einer Standardtumormeldung abgebildet werden. Da die Wohnort(-anamnese), die Familienanamnese, die Tätigkeitsanamnese und der Raucherstatus nur den Zustand zum Zeitpunkt der Meldung und nicht unbedingt den Zustand zum Zeitpunkt Diagnose widerspiegeln.

Relation Standardtumormeldungen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus (Tumor-)Meldungen.
<u>tumormeldung_id</u>	number(1)	Schlüssel aus Tumormeldungen.
diagnose_jahr	number(4)	Vierstelliges Jahr – 0 für unbekannt.
diagnose_monat	number(2)	Ein- bzw. zweistelliger Monat (1-12, 0 für unbekannt).
diagnose_anlass	number(1)	- 1 (Beschwerden) - 2 (Früherkennung) - 3 (Arbeitsmedizinische Untersuchung) - 4 (Nachsorge-Untersuchung) - 5 (Zufallsbefund) - 6 (Zufallsbefund bei Autopsie) - 7 (Sonstiges) - 9 (fehlende Angabe / unbekannt)
diagnose	char(6)	Diagnose nach ICD mit 1 Nachkommastelle (z.B. „unbekannt“ nach ICD-9: 239.9 oder ICD-10: D48.9). Dreisteller i.a. als xxx.9 bzw. C/Dxx.9.
diagnose_version	number(2)	ICD-Version f. Diagnose (9, 10, ...).
histologie	number(4)	Histologie nach ICD-O (z.B. „unbekannt“ nach ICD-O, 2. Auflage: 8000).
histologie_version	char(2)	ICD-O-Version f. Histologie, hierbei vorgestelltes D für die deutsche Ausgabe, also 1, 2, D1, D2.

lokalisierung	char(6)	Lokalisation nach dem Tumorlokalisierungsschlüssel zur ICD-O mit mindestens 1 Nachkommastelle (z.B. „unbekannt“ nach ICD-O, 2. Auflage – entsprechend Version 4 des Lokalisierungsschlüssels: C80.9).
lokalisierung_version	number(2)	Version des Lokalisierungsschlüssels (... , 3, 4, ...).
dignitaet	number(1)	Malignitätsgrad aus dem ICD-O-Morphologie-Schlüssel (unabhängig von obigen Versionen): - 0 (gutartig) - 1 (unbest. Charakter / unsicher ob bös- oder gutartig) - 2 (in Situ) - 3 (bösartig / Primärsitz) - 6 (bösartig / Metastase) - 9 (fehl. Angabe / bösartig, aber unbek. ob Metastase)
seite	number(1)	- 1 (rechts) - 2 (links) - 3 (beidseits) - 8 (trifft nicht zu) - 9 (fehlende Angabe / unbekannt)
diagnosesicherung	number(1)	Beste Sicherung der Diagnose: - 1 (klinisch) - 2 (spezielle Diagnostik) - 3 (zytologisch) - 4 (histologisch) - 5 (autoptisch) - 6 (Sonstiges) - 9 (fehlende Angabe / unbekannt)
ausbreitung	number(1)	- 0 (in Situ) - 1 (lokal begrenzt) - 2 (regionäre Lymphknoten / Nachbarschaft) - 3 (Fernmetastasen) - 4 (Systemerkrankung) - 9 (fehlende Angabe / unbekannt)
grading	number(2)	Differenzierungsgrad: - 1 (gut differenziert (G1)) - 2 (mäßig differenziert (G2)) - 3 (schlecht differenziert (G3)) - 4 (undifferenziert (G4)) - 5 (T-Zell-Lymphom) - 6 (B-Zell-Lymphom) - 7 (Null-Zell-Lymphom) - 9 (fehlende Angabe / n. bestimmbar / n. zutreffend) - 16 (Low grade (G1 oder G2)) - 17 (Medium grade (G2 oder G3)) - 18 (High grade (G3 oder G4)) - 19 (Grenzfall bzw. Borderline – nur bei Ovar)
tumorfolge	number(1)	- 1 (erster Tumor) - 2 (zweiter Tumor) - 3 (weiterer Tumor) - 5 (früheres Tumorleiden vorhanden) - 6 (Metastase, Primärtumor unbekannt) - 9 (fehlende Angabe / unbekannt)

Das *Diagnosedatum* gibt den Zeitpunkt, zu dem die gemeldete Tumorerkrankung erstmals ärztlich diagnostiziert bzw. eine Verdachtsdiagnose geäußert wurde, an.

In der Relation *Diagnose_Texte* findet sich ein Klartext zur Diagnose, und in der Relation *TNM_Texte* sind ggf. Angaben zum TNM-Code.

Das Register soll das Auftreten bösartiger Neubildungen einschließlich ihrer Frühstadien erheben. Während der Erprobungsphase werden alle eingehenden Meldungen übernommen und geprüft, ob sie diese Vorgabe erfüllen. Die Behandlung von Tumoren mit unbestimmten Verhalten (ICD-9 235 bis 239 bzw. ICD-10 D37 bis D48), ist noch zu diskutieren. Allen Fällen, in denen gutartige Tumore gemeldet werden, wird nachgegangen, um zu klären, wie es zu diesen Meldungen kommen konnte und um Maßnahmen zu entwickeln, die derartige Meldungen im Routinebetrieb wirksam verhindern können.

Über die Kodierung von Metastasen ist noch keine abschließende Einigung erzielt. Während in der Registerstelle nach dem Abgleich und Best-of durch geeignete Umkodierung¹⁰ keine Metastasen mehr explizit in der Datenbank repräsentiert sein sollen, bleibt es vorerst den Meldern noch vorbehalten, Metastasen direkt als Fälle zu melden.

Werden zum *Grading* vom Melder Intervalle (z.B. G1-G2) angegeben, so soll mittelfristig gemäß [EWLP95] nur der höhere Grad gespeichert werden. Insbesondere ist die Verwendung von Low bzw. High Grade nur dann vorgesehen, wenn dies auch explizit gemeldet wird; Medium Grade wird dann nicht mehr belegt. Zur Evaluation der eingehenden Meldungen werden jedoch in der Anfangsphase des Registerbetriebs entsprechende Intervallangaben stets auf Low, Medium oder High Grade umkodiert.

Alle Melder sind angehalten, aus Angaben über frühere Tumoren – soweit deren Vollständigkeit und Verlässlichkeit eine sinnvolle Dokumentation erlaubt – separate Meldungen zu generieren. Klartexte zu früheren Tumoren werden nur in Ausnahmefällen im Zuge der Meldungsbearbeitung in der Vertrauensstelle zur Definition eigener Meldungen verwendet und gehen ansonsten nicht als eigenständige Tumoren in den durch die Registerstelle ausgewerteten Datenbestand ein. Wenn ein derartiger Klartext in der Relation *Fruehere_Tumoren_Texte* vorhanden ist, darf unter *Tumorfolge* weder „erster Tumor“ noch „fehlende Angabe“ spezifiziert sein.

Relation Pathologenmeldungen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
geschlecht	number(1)	<s. Standardmeldungen>
geburtsjahr	number(4)	<s. Standardmeldungen>
wohnt	number(10)	<s. Standardmeldungen>, vorerst Kodierung des Landkreises, später (für den Routinebetrieb) regionale Beobachtungseinheit.

Für die Umsetzung des Meldeweges für Pathologen (periphere Teilanonymisierung mit zentraler Verschlüsselung) muß zumindest der *Wohnort*, also die regionale Beobachtungseinheit, in der Vertrauensstelle gespeichert werden, da er erst dort aus der Anschrift generiert wird. Die Attribute *Geschlecht*, *Geburtsjahr* und *Wohnort* werden ebenfalls aus den vorgesehenen personenidentifizierenden Attributen abgeleitet. Tumorspezifische Daten, wie sie in der Registerstelle in der Relation *Pathologentumormeldung* geführt werden (s. Abschnitt 2.2.3), liegen in der Vertrauensstelle nicht vor.

¹⁰ Die *Lokalisation* der Metastase könnte erhalten bleiben; die *Diagnose* wird zu einem unbekanntem Primärtumor.

Relation Sterbemeldungen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
arzt_ref	number(5)	Fremdschlüssel aus Ärzte – Verweis auf den Haus- oder behandelnden Arzt, ggf. NULL.
geschlecht	number(1)	<s. Standardmeldungen>
geburtsjahr	number(4)	<s. Standardmeldungen>
geburtsmonat	number(2)	<s. Standardmeldungen>
staatsangehoerigkeit	number(3)	<s. Standardmeldungen>
staat_version	number(4)	<s. Standardmeldungen>
wohntort	number(10)	<s. Standardmeldungen>
geocode_x	number(7)	<s. Standardmeldungen>
geocode_y	number(7)	<s. Standardmeldungen>
todesjahr	number(4)	Vierstelliges Jahr – 0 für unbekannt.
todesmonat	number(2)	Ein- bzw. zweistelliger Monat (1-12, 0 für unbekannt).
unmittelbare_tu_icd	char(6)	Kodierung der unmittelbaren Todesursache nach ICD – ICD-9: 799.9 für „unbekannt“, nach ICD-10: R99.9.
unmittelbare_tu_version	number(2)	ICD-Version zu unmittelbarer Todesursache.
grundleiden_icd	char(6)	Analog amtliche Todesursache nach ICD.
grundleiden_version	number(2)	Analog ICD-Version zu Grundleiden.
autopsie	number(1)	- 1 (Autopsie durchgeführt) - 2 (Autopsie nicht durchgeführt) - 9 (fehlende Angabe / unbekannt)

In der Relation Wohnort_Texte findet sich ggf. ein Klartext zum Wohnort. Außerdem stehen eventuelle Klartexte zur unmittelbaren Todesursache sowie zum Grundleiden in den entsprechenden Klartext-Relationen. Beim Grundleiden findet sich ggf. eine Beschreibung der Epikrise, also der vollständigen Kausalkette der Todesursachen, falls dies erforderlich sein sollte. Klartexte zu früheren Tumoren können in der Relation Fruehere_Tumoren_Texte abgelegt werden.¹¹

Der meldende *Arzt* (aus der gleichnamigen Relation) ist derjenige, der den Totenschein ausfüllt. Zusätzlich zu diesem wird hier ein Verweis auf den Haus- oder behandelnden Arzt (zum Zwecke des Follow-Backs bei DCO-Fällen) abgelegt, der jedoch nicht an die Registerstelle übertragen wird. Dementsprechend muß ein Follow-Back noch vor dem Abschluß der Meldungsbearbeitung und dem damit verbundenen Löschen der Meldung in der Vertrauensstelle (spätestens nach 3 Monaten) initiiert werden.

Sind auf einem Totenschein gleichzeitig auch Angaben zur Autopsie vorhanden, so werden hieraus zwei getrennte Meldungen (einmal *Meldemodus* „Todesbescheinigung“, einmal „Autopsie“) generiert. Der Verlust der Information, daß Autopsie und Totenschein gemeinsam eingegangen sind, wird in Kauf genommen, da dieses Wissen nach den bisherigen Überlegungen für das Register keine Rolle spielt – insbesondere hat es keinen Einfluß auf die Markierung von DCO-Fällen: derartige Fälle sind *keine* DCO- sondern DCN-Fälle.

Das *Grundleiden* ist gleich der amtlichen Todesursache vom Statistischen Landesamt. Diese amtliche Todesursache kann von dem Grundleiden des Totenscheines abweichen, und daher muß die Todesursache des Statistischen Landesamtes über die Sterbebuchnummer des Totenscheines als *Grundleiden* in die Relation Sterbefaelle übernommen werden. Eine Sterbemeldung darf erst

¹¹ Jedoch sollte bei Sterbemeldungen noch stärker als bei Standardmeldungen forciert werden, daß frühere Tumoren bereits durch den Melder als eigene Tumormeldungen kodiert werden – zumal das Schema hierauf ausgerichtet ist und über den Zeitraum seit Erkrankung auch unterschiedliche Diagnosedaten repräsentieren kann.

nach diesem Vorgang an die Registerstelle übermittelt werden. Die unmittelbare Todesursache wird zu Zeit in der Vertrauenstelle nicht erfaßt und der Sinn dieses Attributs ist wegen der relativ geringen Aussagekraft fraglich. Es ist daher möglich, daß dieses Attribut in zukünftigen Versionen des Datenschemas entfernt wird.

Relation Sterbetumormeldungen

Attribut	Typ	Ausprägungen / Kommentar
meldung_id	number(8)	Schlüssel aus (Tumor-)Meldungen.
tumormeldung_id	number(1)	Schlüssel aus Tumormeldungen.
art_der_todesursache	number(1)	- 1 (unmittelbare Todesursache) - 2 (als Folge von) - 3 (Grundleiden) - 4 (weitere wesentliche Erkrankung) - 5 (Epikrise) - 9 (fehlende Angabe / unbekannt)
diagnose_jahr	number(4)	<s. Standardmeldungen>
diagnose_monat	number(2)	<s. Standardmeldungen>
diagnose_anlass	number(1)	<s. Standardtumormeldungen>
diagnose	char(6)	<s. Standardtumormeldungen>
diagnose_version	number(2)	<s. Standardtumormeldungen>
histologie	number(4)	<s. Standardtumormeldungen>
histologie_version	char(2)	<s. Standardtumormeldungen>
lokalisierung	char(6)	<s. Standardtumormeldungen>
lokalisierung_version	number(2)	<s. Standardtumormeldungen>
dignitaet	number(1)	<s. Standardtumormeldungen>
seite	number(1)	<s. Standardtumormeldungen>
diagnosesicherung	number(1)	<s. Standardtumormeldungen>
ausbreitung	number(1)	<s. Standardtumormeldungen>
grading	number(2)	<s. Standardtumormeldungen>
tumorfolge	number(1)	<s. Standardtumormeldungen>
tod_tumorbedingt	number(1)	- 1 (Tod durch diesen Tumor bedingt) - 2 (Tod nicht durch diesen Tumor bedingt) - 3 (fraglich, ob Tod durch diesen Tumor bedingt) - 9 (fehlende Angabe / unbekannt)

In der Relation Diagnose_Texte findet sich ein Klartext zur Diagnose, und in der Relation TNM_Texte sind ggf. Angaben zum TNM-Kode.

Als *Diagnoseanlaß* wird derzeit „fehlende Angabe / unbekannt“ kodiert. Für eine Spezifikation von „Zufallsbefund bei Autopsie“ wäre bisher nicht verfügbares Zusatzwissen (über den Verlauf der Autopsie oder spezielle, nur bei dieser verwendete Kodierungen) nötig.

Als *Diagnosesicherung* bei Totenscheinen wird „unbekannt“ oder – falls eine Autopsie durchgeführt wurde – „autoptisch“ kodiert (in diesem Fall gehen wir davon aus, daß eine histologische Untersuchung vorgenommen wurde).

Das *Diagnosedatum* bei der Sterbetumormeldung ist, der vom Austeller der Meldung angegebene, Erkrankungszeitpunkt. Totenscheinen können beispielsweise diese Angabe enthalten.

1.2.2 Weitere meldungsbezogene Angaben

Relation Taetigkeiten

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>taetigkeit_id</u>	number(1)	Fortlaufende Nummer der Tätigkeiten einer Meldung.
beruf	number(4)	Kodierung der Tätigkeit nach der Klassifikation der Berufe der Bundesanstalt für Arbeit – 9911 für unbekannt.
beruf_version	number(4)	Version der Kodierung (Jahr des Erscheinens der Auflage) – 0 für unbekannt.
taetigkeit_dauer	number(4)	Dauer der Berufsausübung (bis zum Zeitpunkt der Meldung) in Monaten – 0 für unbekannt.
letzte_taetigkeit	number(1)	War diese Tätigkeit die letzte vor der Diagnose des jeweiligen Tumors der zugehörigen Meldung? - 1 (ja) - 2 (nein) - 9 (fehlende Angabe / unbekannt)
laengste_taetigkeit	number(1)	War diese Tätigkeit die längste vor der Diagnose des jeweiligen Tumors der zugehörigen Meldung? - 1 (ja) - 2 (nein) - 9 (fehlende Angabe / unbekannt)

In der Relation Beruf_Texte findet sich ggf. ein nicht standardisierter Klartext zum Beruf.

Die Tätigkeiten auf einer Meldung werden fortlaufend durchnummeriert, d.h. der Primärschlüssel trägt keinerlei Semantik.

Relation Wohnortanamnesen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
wohnung_seit_min	number(4)	Angabe eines Bereichs von Jahreszahlen (Unterstützung unscharfer Angaben, i.a. Min = Max), seit dem der aktuelle Wohnort Gültigkeit hat – 0 für unbekannt.
wohnung_seit_max	number(4)	<s. wohnort_seit_min>
geburtsort_gkz	number(10)	Gemeindekennziffer des Geburtsortes (auch außerhalb Niedersachsens, falls vorhanden) – 0 falls undefiniert (außerhalb Deutschlands) oder unbekannt.
geburtsland	number(3)	Ländercode des Statistischen Bundesamtes für das Geburtsland (möglichst inklusive Kodierung des Bundeslandes in Deutschland) – 999 für unbekannt.
geburtsort_text	char(80)	Klartext für Geburtsort und -land.
aufwuchs_gkz	number(10)	Analog zum Geburtsort Gemeindekennziffer des Ortes des Aufwachsens – 0 falls undefiniert oder unbekannt.
aufwuchs_land	number(3)	Analog zum Geburtsland Ländercode des Statistischen Bundesamtes für den Ort des Aufwachsens – 999 für unbekannt.
aufwuchs_text	char(80)	Klartext für Ort und Land des Aufwachsens.
laengster_aufenthalt_ort	number(10)	Kodierung des Ortes des längsten Aufenthalts analog zum Wohnort in Relation Standardmeldungen.

laengster_aufenthalt_land	number(3)	Analog zum Geburtsland Ländercode des Statistischen Bundesamtes für das Land des bisher überwiegenden Aufenthalts – 999 für unbekannt.
laengster_aufenthalt_text	char(80)	Klartext für Ort und Land des bisher überwiegenden Aufenthalts (inklusive Anschrift).
staat_version	number(4)	Jeweils verwendete Version der Länderkodierungen – 0 für unbekannt.

Zu jeder Meldung gibt es maximal eine Wohnortanamnese.

Alle Ortsangaben erfolgen nach aktuellen Grenzen zum Zeitpunkt der Meldungsbearbeitung. Die Gemeindegrenznummern sollen so genau wie möglich angegeben werden, insbesondere außerhalb von Niedersachsen können aber auch gröbere Aggregierungsebenen (Landkreise, Regierungsbezirke, Bundesländer) verwendet werden.

Der *längste Aufenthaltsort* soll – analog zum Wohnort zum Diagnosezeitpunkt – zukünftig nicht immer als Gemeindegrenznummer, sondern in den Fällen, in denen große Gemeinden in regionale Beobachtungseinheiten feiner untergliedert werden, durch solche kodiert werden. Im Klartextfeld wird die hierzu benötigte Anschrift nach der Bildung der Beobachtungseinheit für den Transfer in die Registerstelle gelöscht.

Relation Familienanamnesen

Attribut	Typ	Ausprägungen / Kommentare
meldung_id	number(8)	Schlüssel aus Meldungen.
familie_id	number(1)	Fortlaufende Nummer der Tumoren in der Familie auf einer Meldung.
verwandschaft	number(1)	Grad der Verwandtschaft: - 0 (Enkel) - 1 (Kinder) - 2 (Geschwister) - 3 (Eltern) - 4 (Neffen) - 5 (Großeltern) - 6 (Onkel / Tanten) - 7 (Cousins / Cousinen) - 8 (Sonstige) - 9 (fehlende Angabe / unbekannt)
geschlecht	number(1)	<s. Standardmeldungen>
diagnose	char(6)	Diagnose der Krebserkrankung nach ICD.
diagnose_version	number(2)	<s. Standardtumormeldungen>

Die Angaben zu Tumoren in der Familie auf einer Meldung werden fortlaufend durchnummeriert, d.h. der Primärschlüssel trägt keinerlei Semantik.

Es können zu jeder Meldung beliebig viele Einträge erfolgen – auch mit gleichem Verwandtschaftsgrad. Es ist nicht festzustellen, ob sich zwei Einträge mit gleichem Verwandtschaftsgrad und gleichem Geschlecht auf dieselbe Person beziehen. (Dies wird im Hinblick auf die angestrebte Nutzung und Auswertung der Angaben zur Familienanamnese als nicht erforderlich erachtet.)

Relation Therapien

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus (Tumor-)Meldungen.
<u>tumormeldung_id</u>	number(1)	Schlüssel aus Tumormeldungen.
<u>therapie_id</u>	number(1)	Fortlaufende Nummer der Therapien zu einem Tumor auf einer Meldung.
therapieart	number(1)	- 1 (Operation) - 2 (Radiatio) - 3 (Chemotherapie) - 4 (Hormontherapie) - 5 (Immuntherapie) - 6 (Knochenmarkstransplantation) - 7 (Sonstige) - 9 (fehlende Angabe / unbekannt)
therapieziel	number(1)	- 1 (kurativ) - 2 (palliativ) - 3 (adjuvant) - 4 (supportiv) - 5 (neoadjuvant) - 6 (explorativ) - 7 (Sonstige) - 9 (fehlende Angabe / unbekannt)
therapiestatus	number(1)	- 1 (durchgeführt) - 2 (nicht durchgeführt) - 3 (vorgesehen) - 4 (verweigert)

Angaben zu Therapien eines Tumors auf einer Meldung werden fortlaufend durchnummeriert. Es soll lediglich die Primärtherapie erfaßt werden.

Zu jedem Tumor können mehrere Einträge erfolgen, die jeweils unterschiedliche Aussagen über Art und Ziel von Einzelmaßnahmen machen.

Relation Standesamtinfos

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Sterbemeldungen.
standesamtnr	number(8)	Schlüssel aus K_Standesaemter.
sterbebuchnr	number(4)	Sterbebuchnummer in dem obigen Standesamt.
jahrgang	number(4)	Sterbejahrgang.

Jedes Standesamt erfaßt alle Todesfälle in seinem Einzugsbereich nach dem Sterbejahrgang und einer laufenden Sterbebuchnummer. Zu jeder Sterbemeldung, die von einem Totenschein eines Gesundheitsamts stammt, gibt es genau ein Tupel in dieser Relation.

1.2.3 Klartexte

An verschiedener Stelle werden in den Meldungen Klartexte erfaßt, die die vorhandenen Kodierungen ergänzen bzw. für die spätere Überprüfung oder Revision konkretisieren. Diese Texte werden in gesonderten Relationen abgelegt. Im einzelnen ergeben sich die Relationen

- Wohnort_Texte (zu Standard- und Sterbemeldungen) – auch hierauf kann (zumindest sobald jeweils exakte Gauß-Krüger-Koordinaten vorliegen) eventuell verzichtet werden -,
- Todesursache- und Grundleiden_Texte (jeweils zu Sterbemeldungen).
- Fruehere_Tumoren_Texte (zu Standard- und Sterbemeldungen – jeweils ein Klartext mit Angaben zu beliebig vielen früheren Tumoren),
- Diagnose_Texte (zu Standard- und Sterbetumormeldungen inkl. anderer Stadienklassifikationen als TNM sowie beliebiger Bemerkungen auf Erhebungsbögen),
- Beruf_Texte (zu Tätigkeiten) – hier ist noch nicht endgültig entschieden, ob Berufe überhaupt im Klartext oder nur kodiert erfaßt werden sollen -,
- TNM_Texte (zu Standard- und Sterbetumormeldungen),

Ein Klartext ist jeweils in maximal 999 Blöcke mit je maximal 80 Zeichen aufgesplittet.¹² Die Relationen haben somit folgende Form:

Relationen Wohnort-, Todesursache-, Grundleiden- und Fruehere_Tumoren_Texte

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>zeile</u>	number(3)	Blöcke eines Klartextes ab 1.
text	char(80)	Der Text.

Relation Diagnose_Texte

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus (Tumor-)Meldungen.
<u>tumormeldung_id</u>	number(1)	Schlüssel aus Tumormeldungen.
<u>zeile</u>	number(3)	Blöcke eines Klartextes ab 1.
text	char(80)	Der Text.

Relation Beruf_Texte

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen / Tätigkeiten.
<u>taetigkeit_id</u>	number(1)	Schlüssel aus Tätigkeiten.
<u>zeile</u>	number(3)	Blöcke eines Klartextes ab 1.
text	char(80)	Der Text.

Eine Ausnahme von den übrigen Klartexttabellen bilden Angaben zum TNM-Kode (vgl. Anhang). Hier wird der Tatsache Rechnung getragen, daß auf einer Meldung mehrere TNM-

¹² Eventuell soll diese blockweise Speicherung später einmal durch die Verwendung von *long*-Feldern in ORACLE ersetzt werden.

Kodierungen (z.B. sowohl klinisch als auch auf Basis einer histologischen Untersuchung) angegeben werden können. Da zudem das Format der Codes sehr frei ist, wird hierfür eine separate Relation geführt, die neben dem Klartext auch daraus extrahierte einfache Codes anbietet: Diese einfacheren Codes entstehen dadurch, daß man sich auf jeweils auf die Kategorien (z.B. T1) von T, N und M beschränkt und auf eine weitere Differenzierung (z.B. T1a) verzichtet. Die TNM-Version hat primär für den TNM-Klartext eine Bedeutung. Die Attribute T, N und M, die die jeweiligen Kategorien beinhalten, sind in dieser undifferenzierteren Kodierung wesentlich unabhängiger von der TNM-Version. Eine Umkodierung einer älteren TNM-Version in eine neuere Version ist auf Grund fehlender Angaben im allgemeinen nicht möglich.

Relation TNM_Texte

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus (Tumor-)Meldungen.
<u>tumormeldung_id</u>	number(1)	Schlüssel aus Tumormeldungen.
<u>tnm_id</u>	number(1)	Fortlaufende Numerierung der TNM-Angaben einer Tumormeldung – die Reihenfolge trägt keine Semantik.
tnm_klartext	char(30)	Der TNM-Kode inkl. C-Faktoren als Klartext aus der Meldung (etwa in der Form ypT1(m)C1cN1C1M1C1) – evtl. auch mit Angaben zum Ort von Metastasen.
tnm_art	number(1)	Die Art des TNM-Kodes wird aus dem Klartext extrahiert: - 1 (cTNM) - 2 (pTNM) - 3 (y-cTNM – nach initialer multimodaler Therapie) - 4 (y-pTNM) Die Unterscheidung von cTNM und pTNM bezieht sich im Zweifelsfall auf die Klassifikation des Tumors.
t	char(2)	Tumorausbreitung. Ausprägungen: is, a, 0, 1, 2, 3, 4 und X für unbekannt.
ct	number(1)	C-Faktor der Tumorausbreitung. Ähnlich der Diagnosesicherung: - 1 (klinisch) - 2 (spez. Diagnostik, evtl. auch Biopsie/Zytologie) - 3 (chirurg. Exploration mit Biopsie/Zytologie) - 4 (Tumorsektion mit patholog. Untersuchung) - 5 (autoptisch) - 9 (fehlende Angabe / unbekannt)
n	char(1)	Ausbreitung in die Lymphknoten. Ausprägungen: 0, 1, 2, 3 und X für unbekannt.
m	char(1)	Vorliegen von Fernmetastasen. Ausprägungen: 0, 1 und X für unbekannt.
tnm_version	number(2)	Version und Auflage des TNM-Schlüssels (etwa 41 für 4. Version, 1. Auflage oder 40 für 4. Version, unbekannte Auflage) für Angaben unter T, N – M sollte unabhängig von der Version sein – 0 für unbekannt.

Weiterhin kann zu jeder Meldung und Tumormeldung bei der Bearbeitung ein Kommentar gespeichert werden, der z.B. ungewöhnliche Merkmale bzw. Merkmalskombinationen erläutert oder die Ergebnisse von Nachfragen schildert und jeweils mit der Meldung übertragen wird:

Kommentar_Meldung

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>zeile</u>	number(3)	Blöcke eines Klartextes ab 1.
text	char(80)	Der Text.

Kommentar_Tumormeldung

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>tumormeldung_id</u>	number(1)	Schlüssel aus Tumormeldungen.
<u>zeile</u>	number(3)	Blöcke eines Klartextes ab 1.
text	char(80)	Der Text.

1.2.4 Personenidentifizierende Daten, Kontrollnummern und Chifftrate

In der Vertrauensstelle werden personenidentifizierende Daten gespeichert, um daraus Kontrollnummern zu bilden, und – außer bei Pathologenmeldungen – anschließend chiffriert an die Registerstelle geschickt:

Relation Personenidentifizierende_Daten

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
name_1	char(40)	1. Teil des Nachnamens.
name_2	char(40)	2. Teil des Nachnamens.
name_3	char(40)	3. Teil des Nachnamens.
vorname_1	char(40)	1. Teil des Vornamens.
vorname_2	char(40)	2. Teil des Vornamens.
vorname_3	char(40)	3. Teil des Vornamens.
geburtsname_1	char(40)	1. Teil des Geburtsnamens.
geburtsname_2	char(40)	2. Teil des Geburtsnamens.
geburtsname_3	char(40)	3. Teil des Geburtsnamens.
frueherer_name_1	char(40)	1. Teil eines früheren Namens.
frueherer_name_2	char(40)	2. Teil eines früheren Namens.
frueherer_name_3	char(40)	3. Teil eines früheren Namens.
geschlecht	number(1)	- 1 (männlich) - 2 (weiblich) - 9 (fehlende Angabe/ unbekannt).
geburtsdatum	date	
strasse	char(30)	
hausnummer	char(6)	
plz	number(5)	Postleitzahl des Wohnorts
ort	char(30)	
sterbedatum	date	
diagnosedatum	date	
titel_1	char(10)	1. Teil des Titels.
titel_2	char(10)	2. Teil des Titels.

Es sind – bis auf *Geschlecht* – überall NULL-Werte erlaubt. Eine zu große Zahl von NULL-Werten führt jedoch in der Registerstelle zu Problemen beim Abgleich über die Kontrollnummern und somit zu Rückfragen.

Einige Felder sind redundant zu den epidemiologischen Daten in den Meldungsrelationen (*Geschlecht*, Jahr und Monat von *Geburts-*, *Sterbe-* und *Diagnosedatum* sowie in gewissem Rahmen die Angaben zum Wohnort).

Bei unbekanntem *Geburts-* oder *früheren Namen* bleibt dieser unbekannt und wird nicht etwa durch den aktuellen Namen definiert.

Name, *Vorname*, *Geburtsname* und *früherer Name* werden jeweils in drei Komponenten zerlegt. Die Zerlegung der Namen vereinfacht die Zuordnungen, wenn z. B. nicht immer alle Vornamen angegeben, von einem Doppelnamen nicht alle Teile vorhanden sind oder Namenserverweiterungen („von“, „zu“ usw.) vorkommen. Namenserverweiterungen werden stets in der dritten Komponente abgelegt (vgl. [AFH+96, AFK+95]).

Um Probleme mit unterschiedlichen Darstellungen von Umlauten zu vermeiden, werden alle Namenskomponenten standardisiert, d.h. Umlaute und ß werden umgesetzt (ß→ss, ä→ae, ö→oe, ü→ue) und alles einheitlich groß geschrieben. Ebenfalls wird ein vorhandener Titel in seine Bestandteile zerlegt und in die beiden Einzelattribute abgespeichert (z.B.: Prof. Dr. -> titel_1 := Prof. und titel_2 := Dr.).

Relation Chiffretexte

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>zeile</u>	number(3)	Fortlaufende Numerierung der Textzeilen ab 1.
schlüsseltext	char(80)	Schlüsseltext der RSA-Verschlüsselung der PID.

Der *Schlüsseltext* einer Meldung wird in Zeilen mit jeweils (abgesehen von der letzten Zeile) 80 Zeichen aufgeteilt, die getrennt in einzelnen Tupeln gespeichert werden.

Relation Kontrollnummern

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>kn_id</u>	number(2)	Index der Kontrollnummer (s.u.: 1, ..., 22).
wert	char(23)	Textblock zur Kontrollnummer.

Zu jeder Meldung werden folgende 22 Kontrollnummern gebildet:

kn_index	Kontrollnummer	Bemerkungen
1	Name (1. Teil)	<aus Personenidentifizierende_Daten>
2	Name (2. Teil)	<dito>
3	Name (3. Teil)	<dito>
4	Vorname (1. Teil)	<dito>
5	Vorname (2. Teil)	<dito>
6	Vorname (3. Teil)	<dito>
7	Geburtsname (1. Teil)	<dito>
8	Geburtsname (2. Teil)	<dito>

9	Geburtsname (3. Teil)	<dito>
10	Früherer Name (1. Teil)	<dito>
11	Früherer Name (2. Teil)	<dito>
12	Früherer Name (3. Teil)	<dito>
13	Geburtstag	<dito>
14	DDR-Namenskod	4 Ziffern, wobei die ersten beiden für den Namen und die nächsten beiden für den Vornamen stehen.
15	Phonetischer Kod	Kölner Phonetik des standardisierten Namens.
16	Phonetischer Kod	Kölner Phonetik des standardisierten Vornamens.
17	Phonetischer Kod	Kölner Phonetik des standardisierten Geburtsnamens.
18	Phonetischer Kod	Kölner Phonetik des standardisierten früheren Namens.
19	Titel (1. Teil)	<aus Personenidentifizierende_Daten>
20	Titel (2. Teil)	<dito>
21	1. KN aus Baden-Württemberg	Vorname (3), Name (3), Geburtsdatum.
22	2. KN aus Baden-Württemberg	Vorname (3), Geburtsname (3), Geburtsdatum.

Als Phonetikfunktion wird die Kölner Phonetik verwendet, wobei vor deren Anwendung die drei standardisierten Namenskomponenten aneinandergehängt werden.

Der DDR-Namenskod wird aufgenommen, um Abgleiche mit dem Altbestand des Gemeinsamen Krebsregisters der neuen Bundesländer zu ermöglichen, in dem zum Teil dieser Kod anstelle von Name und Vorname gespeichert ist. Der DDR-Namenskod besteht aus vier Ziffern, wobei die ersten beiden für den Namen und die nächsten beiden für den Vornamen stehen. Dabei werden die Namen in 100 etwa gleich große Klassen eingeordnet. Die Zuordnung entspricht genau dem bei [KW75] beschriebenen Schlüssel des Statistischen Bundesamts.

Die beiden Kontrollnummern aus Baden-Württemberg werden aufgenommen, damit überhaupt ein Abgleich mit Daten aus Baden-Württemberg möglich ist, das laut Landesgesetz die oben angegebenen Kontrollnummern verwenden muß. Bei der Bildung dieser beiden Kontrollnummern sind spezielle Standardisierungen zu berücksichtigen, die von den zuvor genannten abweichen.

Zusätzlich zu den Kontrollnummern werden noch die epidemiologischen Merkmale Geschlecht, Geburtsmonat und -jahr sowie die Gemeindekennziffer zum Record Linkage verwendet.

1.2.5 Kodierungstabellen für Attribute, Dokumentationskräfte und Melder

Zu allen Attributen vom Aufzählungstyp (numerische, ein- oder zweistellige Kodierung) wird jeweils eine Kodierungstabelle vorgehalten, die zu jedem Kod einen Klartext sowie einen kurzen Mnemokod (inbs. für die kompakte Darstellung in CARELIS, dem Abgleichsystem der Registerstelle) definiert. (Die 9 – beim Institutionstyp die 99 – steht hier einheitlich für „fehlende Angabe / unbekannt“, sofern eine derartige Ausprägung vorgesehen ist.) Diese Tabelle könnte ggf. noch erweitert werden um Kurz- oder Langversionen des Textes (z.B. für Beschriftungen in der Benutzungsoberfläche des Auswertungssystems) oder andere numerische oder mnemotechnische Kodes, die den Datentransfer von und zu anderen Systemen erleichtern. Es ergeben sich Relationen als Kodierungen der Attribute

- Geschlecht,

- verstorben,
- Mehrling,
- letzte / längste Tätigkeit,
- Verwandtschaft,
- Tod tumorbedingt,
- Autopsie,
- Raucher-Status,
- Diagnose-Anlaß,
- Dignität,
- Seite,
- Diagnosesicherung,
- CT (Certainty-Factor aus dem TNM),
- Ausbreitung,
- Grading,
- Tumorfolge,
- DCO,
- Therapieart,
- Therapieziel,
- Therapiestatus,
- Meldungsstatus und Validität,
- Patienten-Bemerkung,
- Art der Todesursache,
- TNM-Art,
- Fachrichtung und Institutionstyp (vgl. Abschnitt 1.2.5),
- Patient- und Tumor_Unklarheit, Unklarheitsstatus und Kandidatenstatus (vgl. Abschnitt 1.2.7).

Die gleichnamigen Relationen (wobei Leerzeichen und Bindestriche durch einen Unterstrich ersetzt werden und ein „K_“ vorangestellt wird) haben somit jeweils die Form:

Attribut	Typ	Ausprägungen / Kommentar
<u>auspraegung</u>	number(2)	Wie im Schema jeweils angegeben.
<u>kommentar</u>	char(50)	Der Text zur Ausprägung.
<u>mnemocode</u>	char(1)	Ein mnemotechnischer Kode, insb. für die kompakte Darstellung im Abgleichsystem

Falls ein einstelliger *Mnemocode* nicht ausreicht (derzeit nur im Fall der Fachrichtung), wird das Attribut *mnemocode2* (2 Zeichen) verwendet.

Die Kodierungstabelle zum Meldemodus repräsentiert zusätzlich die Zuordnung der Modi zu den verschiedenen Meldungsarten im Datenschema (vgl. Abschnitt 1.1.1):

Relation K_Meldemodus

Attribut	Typ	Ausprägungen / Kommentar
<u>auspraegung</u>	number(2)	Wie im Schema angegeben.
kommentar	char(50)	Der Text zur Ausprägung.
mnemocode	char(1)	Ein mnemotechnischer Kode, insb. für die kompakte Darstellung im Abgleichsystem
meldungstyp	char(20)	„Standard-“, „Pathologen-“, „Sterbe-“ oder „Anfragemeldung“.

Speziell zur Kodierung von Diagnosen (vgl. Anhang) sind folgende Relationen vorgesehen:

Relation K_ICD

Attribut	Typ	Ausprägungen / Kommentar
<u>icd_version</u>	number(2)	Schlüssel aus K_ICD_Version.
<u>icd_auspraegung</u>	char(6)	ICD-Kode (auch dreistellige Codes).
icd_kommentar	char(120)	Der Text zur Ausprägung.

Die View **K_ICD_KREBS** definiert hierzu eine Sicht, die nur die primären bösartigen Neubildungen (ohne Metastasen) bzw. deren Vorstufen (in Situ) und solche unsicheren Verhaltens (siehe auch Seite 11) enthält. Dies sind für die ICD-9 die Codes 140-195, 199-208 und 230-239¹³ sowie für die ICD-10 die Codes C00-C76, C80-C97, D00-D09 und D37-D48 (vgl. Anhang).

Relation K_ICD_Version

Attribut	Typ	Ausprägungen / Kommentar
<u>icd_version</u>	number(2)	ICD-Revision (etwa 9 oder 10).
kommentar	char(50)	Im Klartext, etwa "ICD, 9. Auflage".

Relation K_Histologie

Attribut	Typ	Ausprägungen / Kommentar
<u>histologie_version</u>	char(2)	Schlüssel aus K_Histologie_Version.
<u>histologie_auspraegung</u>	number(4)	Histologiekode.
<u>dignitaet</u>	number(1)	<s. Standardtumormeldungen>
histologie_kommentar	char(120)	Der Text zur Ausprägung.

Relation K_Histologie_Version

Attribut	Typ	Ausprägungen / Kommentar
<u>histologie_version</u>	char(2)	ICD-O-Version zur Histologie, hierbei vorgestelltes D für die deutsche Ausgabe, also etwa 1, 2, D1, D2.

¹³ Derzeit wird auch die 239 hier noch aufgenommen (Tumoren unbekanntes Verhaltens), insb. um einen Vergleich mit der D48 der ICD-10 (Tumoren unbekanntes *oder* unsicheres Verhaltens) anstellen zu können.

kommentar	char(50)	Klartext, etwa "Histologie nach ICD-O, 2. Auflage".
-----------	----------	---

Relation K_Lokalisation

Attribut	Typ	Ausprägungen / Kommentar
<u>lokalisierung_version</u>	number(2)	Schlüssel aus K_Lokalisation_Version.
<u>lokalisierung_auspraegung</u>	char(6)	Lokalisationskode.
lokalisierung_kommentar	char(120)	Der Text zur Ausprägung.

Relation K_Lokalisation_Version

Attribut	Typ	Ausprägungen / Kommentar
<u>lokalisierung_version</u>	number(2)	Auflage des Tumorlokalisierungsschlüssels (etwa 3 oder 4).
kommentar	char(50)	Klartext, etwa "Tumorlokalisierungsschlüssel, 4. Auflage (ICD-O-2)".

Relation K_TNM_Version

Attribut	Typ	Ausprägungen / Kommentar
<u>tnm_version</u>	number(2)	Version und Auflage des TNM-Schlüssels (etwa 41 für 4. Version, 1. Auflage oder 40 für 4. Version, unbekannte Auflage).
kommentar	char(50)	Klartext, etwa "TNM-Schlüssel, 4. Version, 1. Auflage".

Es gibt keine Kodierungstabelle für TNM-Kodes.

Weiterführende Überlegungen zur besseren Berücksichtigung der unterschiedlichen Aggregat-
onsebenen der Kodierungen sind noch anzustellen.

In ähnlicher Form wie für die Diagnosen sind Kodierungstabellen für Berufe und Staaten (vgl.
Anhang) definiert:

Zur Spezifikation der *standardmäßig* im Register (insb. in der Registerstelle für Auswertungen)
verwendeten Versionen dient folgende Tabelle:

Relation K_Versionen

Attribut	Typ	Ausprägungen / Kommentar
<u>icd_version</u>	number(2)	Fremdschlüssel aus K_ICD_Version.
<u>histologie_version</u>	char(2)	Fremdschlüssel aus K_Histologie_Version.

Um einheitliche Auswertungen auf dem Datenbestand der Registerstelle durchzuführen, ist es
notwendig, daß alle Daten mit den gleichen ICD- bzw. Histologie-Versionen kodiert sind. Diese
Versionskennungen werden in genau einem Tupel der Relation K_Versionen gespeichert. Für die
Vertrauensstelle besitzt diese Relation deshalb einen eher informativen Charakter, da dort nicht
direkt genutzt wird. Es ist jedoch sicherlich anzustreben, daß die alle Meldungen auf der Basis
dieser Versionen geschehen, um später Umkodierungen in der Registerstelle zu vermeiden.

Relation K_Beruf

Attribut	Typ	Ausprägungen / Kommentar
<u>beruf_version</u>	number(4)	Schlüssel aus Beruf_Version.
<u>beruf_auspraegung</u>	number(4)	Kode nach der Klassifikation der Berufe der Bundesanstalt für Arbeit.
<u>beruf_index</u>	number(3)	Index der verschiedenen Synonyme oder Untergruppen. Hierbei hat die offizielle Bezeichnung den Index 1.
beruf_kommentar	char(80)	Der Text zur Ausprägung.

Eventuell muß die Menge der Berufe noch – in Absprache mit den Meldern – um Ausprägungen für z.B. Hausfrauen, Rentner, Schüler / Studenten u.ä. erweitert werden.

Relation K_Beruf_Version

Attribut	Typ	Ausprägungen / Kommentar
<u>beruf_version</u>	number(4)	Jahr des Erscheinens der Auflage.
kommentar	char(50)	Klartext, etwa "Klassif. der Berufe der BA f. Arbeit (1995)".

Relation K_Staat

Attribut	Typ	Ausprägungen / Kommentar
<u>staat_version</u>	number(4)	Schlüssel aus Staat_Version.
<u>staat_auspraegung</u>	number(3)	Kode nach dem Staatenschlüssel des Bundesamtes für Statistik.
staat_kommentar	char(80)	Der Text zur Ausprägung.

Relation K_Staat_Version

Attribut	Typ	Ausprägungen / Kommentar
<u>staat_version</u>	number(4)	Jahr des Erscheinens der Auflage.
kommentar	char(50)	Klartext, etwa "Staatenschlüssel des BA für Statistik (1992)".

Wohnorte werden im Rahmen des für die Vertrauensstelle benötigten Ausschnitts des Geoschemas verwaltet, ggf. in ähnlicher Form wie in der Registerstelle (vgl. Abschnitt 4.2.2).

Die Gesundheitsämter geben der Vertrauensstelle auf dem Totenschein die Sterbebuchnummer eines Standesamtes als zusätzliche Information an. Alle existierenden Standesämter sind in der Relation K_Standesaemter vermerkt.

Relation K_Standesaemter

Attribut	Typ	Ausprägungen / Kommentar
standesamtnr	number(8)	Primärschlüssel
standesamt_name	char(40)	Name des Standesamts

Weiterhin gibt es Kodierungstabellen für die Dokumentationskräfte von Vertrauens- und Registerstelle:

Relation VST_Bearbeiter

Attribut	Typ	Ausprägungen / Kommentare
<u>vst_bearbeiter_id</u>	number(2)	Fortlaufende Numerierung ab 0.
name	char(40)	Name / Kürzel der Dokumentationskraft.
db_login	char(20)	Datenbank-Login.

Als spezieller Eintrag mit der ID 0 ist hier „CARMEL“ (entsprechend automatischer Meldungsbearbeitung) kodiert.

Relation RST_Bearbeiter

Attribut	Typ	Ausprägungen / Kommentare
<u>rst_bearbeiter_id</u>	number(2)	Fortlaufende Numerierung ab 1.
name	char(40)	Name / Kürzel der Dokumentationskraft.
db_login	char(20)	Datenbank-Login.

Der Zugriff auf die Dokumentationskräfte der Registerstelle unterstützt in der Vertrauensstelle die Klärung von Unklarheiten (vgl. Abschnitt 1.2.7). Ein spezieller Eintrag „Registerstelle“ repräsentiert hier automatisch erfolgte Bearbeitungen.

Eine Liste möglicher Titel erleichtert die Identifikation von Namenszusätzen für die Zerlegung der personenidentifizierenden Daten:

Relation K_Titel

Attribut	Typ	Ausprägungen / Kommentare
<u>titel_id</u>	number(2)	Fortlaufende Numerierung ab 1.
titel_text	char(10)	Text.

Schließlich werden auch mögliche Melder in Kodierungstabellen verwaltet, d.h. dieser Datenbestand wird im Grunde genommen unabhängig von konkreten Meldungen geführt und sowohl durch Absprache mit potentiellen Meldern als auch aufgrund von Angaben auf Meldungen ergänzt. Soweit nicht anders angegeben, sind jeweils auch NULL-Werte erlaubt (aber i.a. sicher nicht wünschenswert).

Relation Aerzte

Attribut	Typ	Ausprägungen / Kommentar
<u>arzt_id</u>	number(5)	Fortlaufende Nummer.
institution_ref	number(5)	Fremdschlüssel aus Institutionen.
name	char(40)	Immer definiert.
vorname	char(30)	
titel	char(15)	
geschlecht	number(1)	<s. Standardmeldungen>, immer definiert.

fachrichtung	number(2)	Immer definiert (in Anlehnung an die Facharztbezeichnungen nach der Weiterbildungsordnung der Ärztekammer Niedersachsen): - 1 (niedergelassener Praktischer Arzt, Stationsarzt ohne Facharztbezeichnung) - 2 (Allgemeinmedizin) - 3 (Anästhesiologie) - 4 (Arbeitsmedizin) - 5 (Augenheilkunde) - 6 (Chirurgie) - 7 (Frauenheilkunde und Geburtshilfe) - 8 (Haut- und Geschlechtskrankheiten) - 9 (Hals-Nasen-Ohrenheilkunde) - 10 (Innere Medizin: Hämatologie, Internistische Onkologie) - 11 (Innere Medizin (Sonstige)) - 12 (Kinderchirurgie) - 13 (Kinderheilkunde) - 14 (Mund-Kiefer-Gesichts-Chirurgie) - 15 (Neurochirurgie) - 16 (Nuklearmedizin) - 17 (Neurologie, Psychiatrie/Psychotherapie, Nervenheilkunde) - 18 (Orthopädie) - 19 (Pathologie, Neuropathologie) - 20 ((Diagnostische) Radiologie, Strahlentherapie) - 21 (Urologie) - 90 (Zahnmedizin) - 97 (Sonstige (betreuend)) - 98 (Sonstige (nicht betreuend)) - 99 (fehlende Angabe / unbekannt)
bankinstitut	char (40)	
blz	number(8)	
kontonummer	number(10)	
kontoinhaber	char(40)	Name des Kontoinhabers (auch wenn mit Arzt selbst identisch).
telefon	char(20)	
aktENZEICHEN	char(20)	
kv_nummer	char(7)	

Die Liste der *Fachrichtungen* ist noch mit den Möglichkeiten insb. der Sammelmelder abzustimmen, uns entsprechende Angaben machen zu können.

Für eine Abrechnung von Meldungen durch die Vertrauensstelle sind die Angaben zur *Bankverbindung* natürlich erforderlich. Diese können gelöscht werden, falls von einem Melder voraussichtlich keine Meldungen mehr erfolgen.¹⁴

Relation Institutionen

Attribut	Typ	Ausprägungen / Kommentar
<u>institution_id</u>	number(5)	Fortlaufende Nummer.
institutionsname	char(40)	Immer definiert.

¹⁴ Das genaue Vorgehen bei der Vergütung von Meldungen ist noch weitgehend ungeklärt. Insofern können sich hier – auch unter Berücksichtigung verwaltungsrechtlicher Erfordernisse – noch Änderungen ergeben.

institutionstyp	number(1)	Immer definiert: - 1 (Praxis) - 2 (Krankenhaus) - 3 (Institut) - 4 (Gemeinschaftspraxis) - 5 (Sonstige)
abteilung	char(40)	
chefarzt	char(40)	Name des Chefarztes.
strasse	char(30)	
hausnummer	char(6)	
plz	number(5)	
ort	char(30)	

Relation Sammelmelder

Attribut	Typ	Ausprägungen / Kommentar
<u>sammelmelder_id</u>	number(3)	Fortlaufende Nummer.
institutionsname	char(40)	Name des Sammelmelders – immer definiert.
abteilung	char(40)	
strasse	char(30)	
hausnummer	char(6)	
plz	number(5)	
ort	char(30)	
bankinstitut	char(40)	
blz	number(8)	
kontonummer	number(10)	
kontoinhaber	char(40)	Name des Kontoinhabers (auch wenn mit Institutionsname selbst identisch).
ansprechpartner_name	char(40)	Name des Ansprechpartners.
vorname	char(30)	Vorname des Ansprechpartners.
titel	char(15)	Titel desselben.
geschlecht	number(1)	<s. Standardmeldungen>, Geschlecht zum Ansprechpartner – NULL nur, falls dieser undefiniert ist.
telefon	char(20)	Telefonnummer des Ansprechpartners.

1.2.6 Hilfsrelationen

Temporär werden berechnete Gemeindecennenziffern und Gauß-Krüger-Koordinaten zu den Wohnorten der Patienten in einer separaten Relation abgelegt, bevor sie unter den jeweiligen Meldungen selbst gespeichert werden. Außer dem Schlüssel dürfen alle Attribute (zwischenzeitlich) auch NULL sein.

Relation Geocodes

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
wohntort	number(10)	<s. Standardmeldungen>
geocode_x	number(7)	<s. Standardmeldungen>
geocode_y	number(7)	<s. Standardmeldungen>

Die Vorgabewerte des Bearbeitungsprogramms CARTRUST werden für jeden Benutzer in einem Tupel der Relation Voreinstellungen gespeichert. Der Sinn dieser Werte ist es, wiederholte

Eingaben beispielsweise der ICD-Version einzusparen, da viele Meldungen, insbesondere wenn sie von dem gleichen Melder kommen, in den Attributen dieser Relation übereinstimmen.

Relation Voreinstellungen

Attribut	Typ	Ausprägungen / Kommentar
vst_bearbeiter_ref	number(2)	Schlüssel aus VST_Bearbeiter.
eingangsdatum	date	Eingangsdatum.
sammelmelder_ref	number(3)	Schlüssel aus Sammelmelder.
arzt_ref	number(5)	Schlüssel aus Aerzte.
meldemodus	number(2)	Schlüssel aus K_Meldemodus.
icd_version	number(2)	Schlüssel aus K_ICD_Version.
lokalisierung_version	number(2)	Schlüssel aus K_Lokalisation_Version.
histologie_version	char(2)	Schlüssel aus K_Histologie_Version.
staat_version	number(4)	Schlüssel aus K_Staat_Version.
beruf_version	number(4)	Schlüssel aus K_Beruf_Version.
tnm_version	number(2)	Schlüssel aus K_TNM_Version.
standesamtnr	number(8)	Schlüssel aus K_Standesaemter.
sterbebuchjahrgang	number(4)	Sterbebuchjahrgang.

1.2.7 Relationen für Klartextabgleich und Verwaltung

Im folgenden wird auf Relationen eingegangen, die im Zuge von Rückfragen aus der Registerstelle, Auskunftsanforderungen und Einwilligungswiderrufen durch Patienten sowie zur Behandlung von Rückmeldungen des Vitalstatus von Patienten an die Melder benötigt werden (vgl. Abschnitt 2.2.8). Das exakte Vorgehen bei der Bearbeitung derartiger Fälle wird an dieser Stelle nicht beschrieben – hierzu ist der Bericht [HDW97] erschienen.

Relation Meldung_Unklarheiten

Attribut	Typ	Ausprägungen / Kommentar
meldung_id	number(8)	Schlüssel aus Meldungen.
zeile	number(3)	Index des Eintrags zur Meldung.
rst_bearbeiter_ref	number(2)	Der Fragesteller aus der Registerstelle – NULL für Unklarheiten in der Vertrauensstelle.
unklarheitsstatus	number(1)	- 1 (unbearbeitet) - 2 (mit Änderung) - 3 (ohne Änderung – ungeklärt) - 4 (ohne Änderung – bestätigt)
patient_unklarheit	number(1)	Art der Unklarheit: - 1 (Patientenzuordnung unklar) - 2 (Patienten-Best-of unklar) - 3 (Bearbeitung erfolgreich) - 4 (Unklarheit in Vertrauensstelle)
text	char(80)	Textuelle Beschreibung der Unklarheit – ggf. NULL.

Relation Tumormeldung_Unklarheiten

Attribut	Typ	Ausprägungen / Kommentar
meldung_id	number(8)	Schlüssel aus (Tumor-)Meldungen.
tumormeldung_id	number(1)	Schlüssel aus Tumormeldungen.
zeile	number(3)	Index des Eintrags zur Tumormeldung.

rst_bearbeiter_ref	number(2)	Der Anfragersteller aus der Registerstelle – NULL für Unklarheiten in der Vertrauensstelle.
unklarheitsstatus	number(1)	- 1 (unbearbeitet) - 2 (mit Änderung) - 3 (ohne Änderung – ungeklärt) - 4 (ohne Änderung – bestätigt)
tumor_unklarheit	number(1)	Art der Unklarheit: - 1 (Tumorzuordnung unklar) - 2 (Tumor-Best-of unklar) - 3 (Bearbeitung erfolgreich – DCO-Fall) - 4 (Unklarheit in Vertrauensstelle)
text	char(80)	Textuelle Beschreibung der Unklarheit – ggf. NULL.

Falls bereits bei der Meldungsbearbeitung in der Vertrauensstelle oder als Ergebnis der Abgleichsaufbereitung in der Registerstelle Unklarheiten auftreten, wird dies durch Einträge in den beiden obigen Relationen vermerkt.

Die Referenz auf den *Bearbeiter* aus der Registerstelle erleichtert die Kommunikation zwischen Vertrauens- und Registerstelle sowie die Einarbeitung der Nachfrageergebnisse in der Registerstelle.

Der *Status der Unklarheit* beschreibt, ob und in welcher Form die Unklarheit bearbeitet und beseitigt wurde. Er dient auch zur Festlegung der Validität einer Meldung (vgl. Abschnitt 1.2.1).

Über die Art der *Unklarheit* zu Patient oder Tumor werden von der Registerstelle auch der Abschluß der Bearbeitung einer Meldung an die Vertrauensstelle gemeldet sowie auf DCO-Fälle unter den Tumormeldungen hingewiesen.

Verteilt sich der *Text* zur Unklarheit auf mehrere Zeilen, so ist natürlich die Konsistenz der jeweiligen Tupel in den übrigen Attributen zu garantieren.¹⁵ Das Ergebnis der Bearbeitung der Unklarheit wird bei Nachfragen aus der Registerstelle wieder im Textfeld abgelegt und zurück an diese transferiert. Zur dauerhaften Dokumentation von Nachfrageergebnissen und deren Einarbeitung in Vertrauens- oder Registerstelle können ggf. die entsprechenden Kommentar-Relationen (s. Abschnitt 1.2.3) verwendet werden.

Über das Eingangsdatum der Meldung wird die 3-Monatsfrist zur Bearbeitung berücksichtigt und somit eine „Endlosschleife“ von Nachfragen vermieden. Zusätzlich unterstützt dies sicher auch der Unklarheitsstatus.

Relation Anfrage_Kandidaten

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen – die Anfrage- oder Widerspruchsmeldung.
<u>kandidat_index</u>	number(2)	Numerierung der Anfrage-Kandidaten (bei 1 beginnend) nach fallender Sicherheit der Zuordnung – 0, falls kein Kandidat gefunden wurde.
meldung_ref	number(8)	Fremdschlüssel aus Meldungen – der Abgleichskandidat. NULL, falls kein Kandidat gefunden.

¹⁵ Diese Relationen sind nicht in dritter Normalform – somit Gefahr von Update-Anomalien.

kandidatenstatus	number(1)	- 1 (vorläufiger Kandidat) - 2 (Klartextabgleich positiv) - 3 (Klartextabgleich negativ / kein Kandidat gefunden)
------------------	-----------	---

Hier werden zu Auskunfts- und Widerspruchsmeldungen Kandidaten aus CARELIS (Registerstelle) für den Klartextabgleich in der Vertrauensstelle gekennzeichnet. Als *Status* wird auch das abschließende Ergebnis des Klartextabgleichs abgelegt. Wurde von der Registerstelle kein Kandidat zu einer Anfrage gefunden, wird ein Eintrag mit Fremdschlüssel NULL erzeugt.

Relation Loeschanweisungen

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen – die gelöschte Meldung.

Nach dem Klartextabgleich von Kandidaten zu einer Widerspruchsmeldung werden hier die in der Registerstelle zu löschenden Meldungen aufgelistet – im Anschluß an die Übertragung der Anweisung an die Registerstelle wird diese hier wieder gelöscht. Nach dem Löschen muß ggf. ein neues Best-of in der Registerstelle ermittelt werden.

Relation Anfragen

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel gemäß Meldungen – die Anfrage- oder Widerspruchsmeldung. Hier ist also keine referentielle Integrität gewahrt (kein Fremdschlüssel!).
<u>auskunft_id</u>	number(2)	Fortlaufende Numerierung der betroffenen Meldungen zu einer Anfrage – 0, falls keine passenden Meldungen gefunden wurden.
meldung_ref	number(8)	„Fremdschlüssel“ aus Meldungen – die gelöschte oder herausgegebene Meldung. NULL, falls keine passende Meldung gefunden. Wiederum keine referentielle Integrität.
meldemodus	number(2)	Modus der Anfrage- / Widerspruchsmeldung: - 10 (Widerspruchsmeldung) - 11 (Auskunftsanforderung Patient) - 12 (Auskunftsanforderung Forschung). Die übrigen Ausprägungen aus Meldungen werden hier nicht benötigt.
bearbeitungsdatum	date	Datum der Auskunft, des Löschens bzw. der abschließenden Bearbeitung, falls keine Meldungen zur Anfrage gefunden.

In dieser Relation wird über erfolgte Auskünfte und Löschungen¹⁶ dauerhaft Buch geführt.

¹⁶ Auch bei Löschungen soll eine Auskunft an den Patienten (ggf. über einen Arzt) über den Löschungsvorgang, evtl. auch inklusive der gelöschten Daten, erfolgen.

Relation Arzt_Totenschein_Infos

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>arzt_id</u>	number(5)	Schlüssel aus Ärzten.
externe_referenz	char(15)	<s. Meldungen>

Relation Sammelmelder_Totenschein_Infos

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>sammelmelder_id</u>	number(3)	Schlüssel aus Sammelmelder.
externe_referenz	char(15)	<s. Meldungen>

Nach dem Abgleich und der Meldungsbearbeitung in der Registerstelle werden hier zu verstorbenen Patienten mit Bezug auf die jeweilige Sterbemeldung über den Tod zu informierende Melder spezifiziert. Zur Unterstützung der Identifikation des jeweiligen Patienten durch den Melder, d.h. zur Vermeidung eines Klartext-Abgleichs beim Melder, wird die von ihm vergebene *externe Referenz* zum Patienten – falls vorhanden¹⁷ – aus den betreffenden Meldungsdaten extrahiert und hier dazugeschrieben. Falls dies nicht möglich ist, werden dem Melder die Klartextangaben des Totenscheins separat zugeleitet; einen Abgleich mit seinem Datenbestand muß er dann selber realisieren. Auch bei Verwendung der externen Referenz sollte sich der Melder in irgendeiner Form vergewissern, ob die vom Krebsregister gelieferte Meldung zu seinem Patienten paßt.

¹⁷ Hierzu ist insbesondere erforderlich, daß der Patient zu Lebzeiten explizit der Speicherung der Referenz im Register zugestimmt hat.

2. Die Registerstelle

In diesem Kapitel wird das epidemiologische Datenschema (Meldungen, Tumoren, Patienten und damit zusammenhängende Daten) der Registerstelle vorgestellt. Separat davon erfolgt die Darstellung des Geo-Schemas der Registerstelle, in dem die nötigen Geodaten (insb. Verwaltungsgebiete, Bevölkerungs-, Standardbevölkerungs- und ATKIS-Daten) modelliert werden, in Kapitel 4.

2.1 Das ER-Schema

Grundlegende Konzeption

Das epidemiologische Datenschema der Registerstelle basiert auf einer 1:n-Zuordnung jeweils zwischen Patient und Tumor sowie zwischen Tumor und Meldung, d.h. ein Patient kann mehrere Tumoren haben, und jeder Tumor kann von verschiedenen Institutionen mehrfach an das Register gemeldet werden. Hieraus ergibt sich eine Hierarchie von personen-, tumor- und meldungsbezogenen Merkmalen. Beim Vorliegen von sich ergänzenden, aber auch eventuell teilweise widersprüchlichen Mehrfachmeldungen zu einem Patienten oder Tumor werden jeweils die personen- bzw. tumorbezogenen Angaben mittels eines „Best-Of-Generators“ als beste mögliche Informationen aus den verschiedenen Meldungen generiert. Nach bestimmten Kriterien werden jeweils die „verlässlichsten“ bzw. spezifischsten Attribute aus den jeweiligen Meldungen ausgewählt. Zum genauen Vorgehen hierbei werden in Abschnitt 2.2 einige Anregungen gegeben; eine detaillierte Spezifikation ist –aufgrund erster Erfahrungen mit transferierten Datensätzen – in dem Bericht [HDW97] einzusehen.

Die Grundlage des Datenschemas bildet die Meldung (s. Abb. 2.1). Diese wird von der Vertrauensstelle mitsamt aller Angaben aus den zugeordneten Relationen Chiffretexte, Tätigkeiten, Familien- und Wohnortanamnesen sowie Therapien übernommen (vgl. Abschnitt 1.1). Chiffretexte von Meldungen zu Patienten, die bereits seit mehr als 75 Jahren verstorben sind oder deren Geburtstag mehr als 130 Jahre zurückliegt, werden später wieder aus der Datenbank entfernt. Kontrollnummern werden in Form chiffrierter Aggregate jeweils aller Kontrollnummern zu einer Meldung dauerhaft abgespeichert. Zur Durchführung des Datensatzabgleichs wird das Aggregat temporär wieder in die einzelnen Kontrollnummern aufgespalten, die dann in einer Hilfsrelation abgelegt werden. Weiterhin werden die Angaben zu Ärzten, Institutionen und Sammelmeldern bei Bedarf aus der Datenbank der Vertrauensstelle aktualisiert.

Konzeptionell werden Anfragemeldungen sowie Inzidenz- und Sterbemeldungen zu Neuerkrankungen bzw. Todesfällen von Krebspatienten unterschieden. Über einen Datensatzabgleich können Meldungen vorhandenen Patienten und Tumoren zugeordnet werden oder zur Aufnahme neuer Fälle in die Datenbank führen. Hierbei wird zwischen dem Abgleich auf Patientenebene (Patientenmatch – Zuordnung von Meldungen zum gleichen Patienten über zugeordnete Kontrollnummern) und dem auf Tumorebene (Tumormatch – Zuordnung von Meldungen zum gleichen Tumor, gemäß [PCF+94] im wesentlichen über den ICD-Kode, aber auch unter Heranzie-

hung weiterer Attribute) unterschieden. Nach dem Abgleich¹⁸ bezieht sich jede Inzidenzmeldung in den meisten Fällen auf genau einen Tumor (der wiederum exakt einem Patienten zugeordnet ist) – zu einem Tumor können natürlich mehrere Meldungen eingegangen sein. In Einzelfällen kann eine Inzidenzmeldung auch Angaben zu mehreren gleichzeitig diagnostizierten Tumoren enthalten.

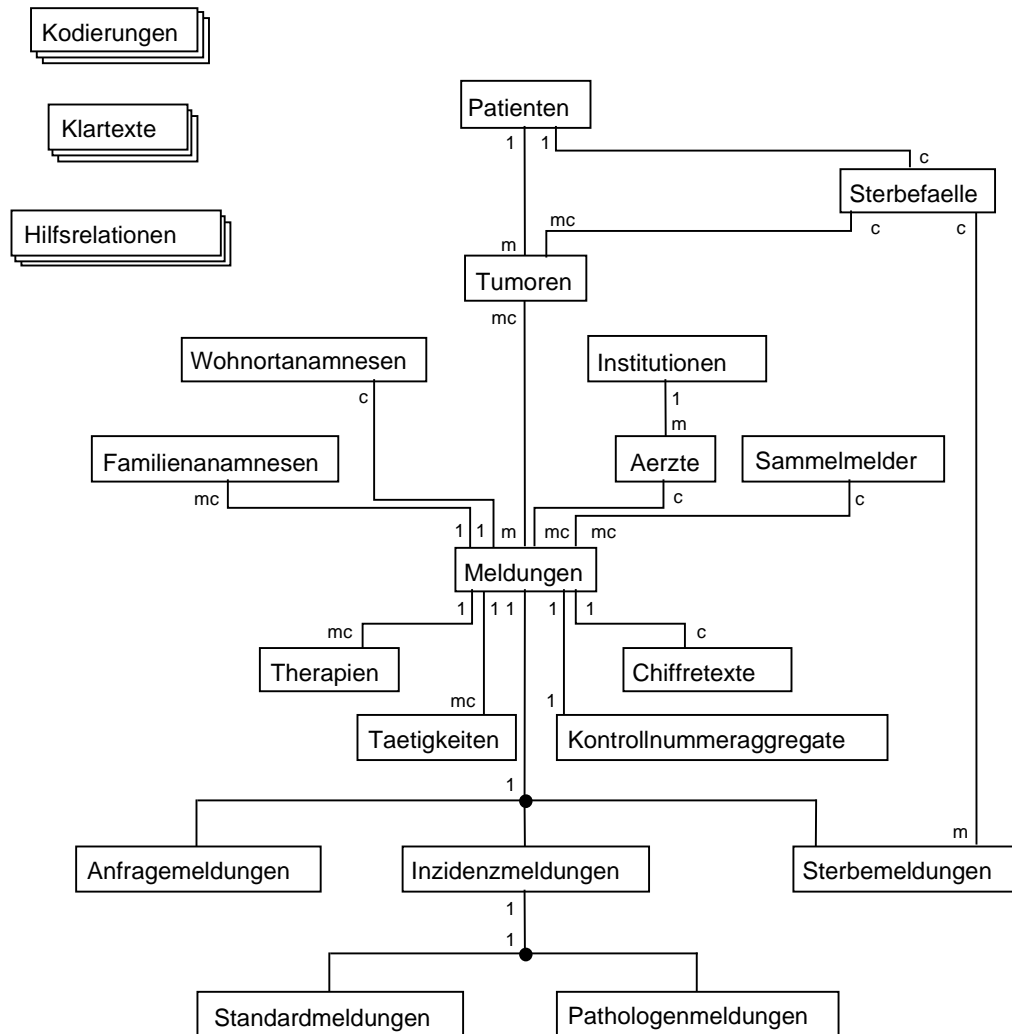


Abb. 2.1: Konzeptionelles ER-Schema der Registerstelle

Jede Sterbemeldung beschreibt (nach dem Abgleich) genau einen, für verstorbene Patienten definierten Sterbefall – zu diesem sind wiederum mehrere Meldungen möglich. Auf einer Sterbemeldung können Angaben zu verschiedenen Tumoren erfolgen (unter unmittelbarer Todesursache, Grunderkrankungen, als Teil der zum Tod führenden Ursachenkette oder unter weiteren Erkrankungen zum Zeitpunkt des Todes). Beim Abgleich wird die Menge der für einen Sterbefall „relevanten“,

¹⁸ Aus der Notwendigkeit, bei der Datenmodellierung sowohl den Zustand vor als auch den nach dem Abgleich zu berücksichtigen, ergibt sich die Vielzahl der c:x- oder mc:x-Beziehungen bei Patient, Sterbefall und Tumor im ER-Schema (Abb. 2.1 und 2.2).

d.h. dem Tod ursächlich zugrundeliegenden Tumoren über eine gesonderte Beziehung zwischen Sterbefall und Tumor definiert.

Klartext- und Kodierungstabellen werden aus dem Schema der Vertrauensstelle übernommen. Hilfsrelationen dienen zur effizienten Arbeit mit den gespeicherten Daten.

Realisierung

Das oben vorgestellte konzeptionelle ER-Schema wird – neben der Auflösung der m:mc-Beziehung zwischen Tumor und Meldung – zur Vereinfachung der Handhabung insbesondere durch Einfügen zusätzlicher Beziehungen leicht modifiziert. Hieraus ergibt sich ein Schema, das dann unmittelbar die Implementierung mittels ORACLE widerspiegelt und im wesentlichen eine Erweiterung des Vertrauensstellenschema (vgl. Abschnitt 1.1) um Patienten- und Tumorbezug darstellt (s. Abb. 2.2). So werden durch die dargestellten Beziehungen direkt die verwendeten Fremdschlüssel definiert, die zu Attributen der jeweils „untergeordneten“ Relation werden.

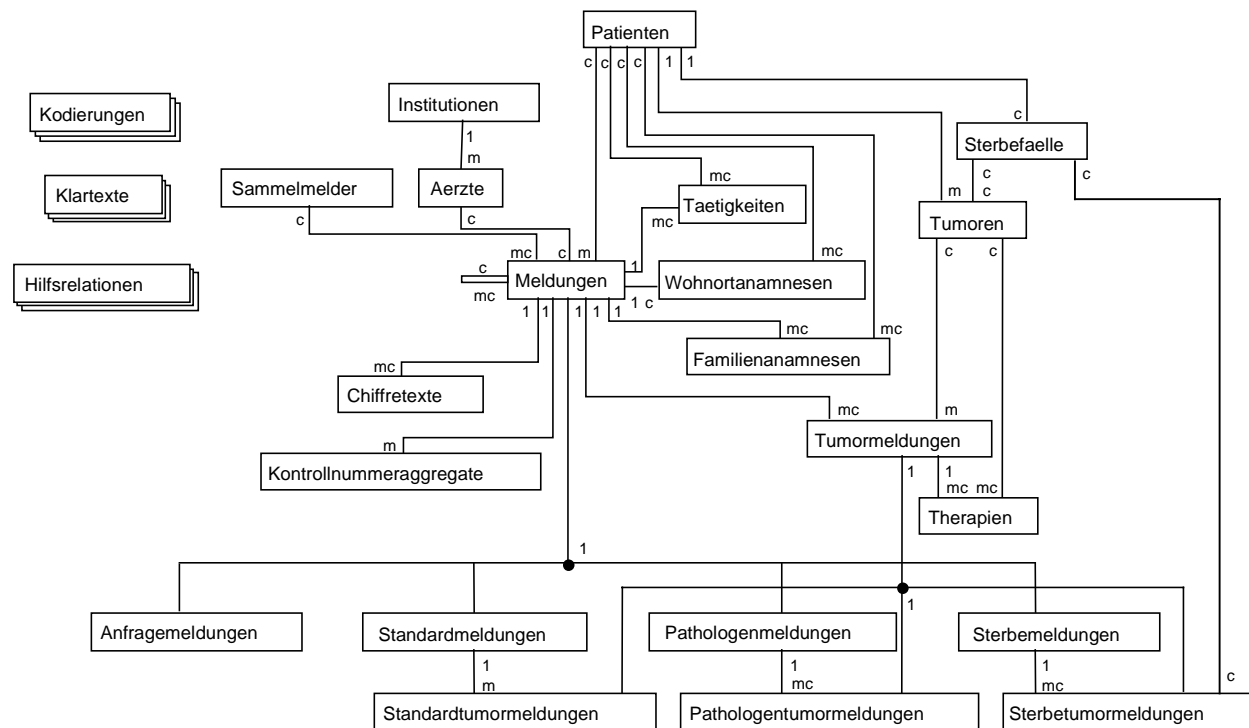


Abb. 2.2: Implementiertes ER-Schema der Registerstelle: Meldungen

Eine wesentliche Änderung besteht in der Aufgabe der Unterscheidung von Inzidenz- und Sterbemeldungen. Mehrere Tumoren auf einer Meldung werden separat als Tumormeldungen verwaltet. Hiermit ergibt sich eine klare Trennung der Informationen zu Tumoren von denen zu Patienten (etwa im Fall von Sterbemeldungen den Angaben zum Tod des Patienten), was auch für den Datensatzabgleich (tumor- bzw. personenbezogen) eine Rolle spielt. Falls gemäß Klartextangaben zu früheren Tumoren aus einer Meldung zusätzliche neue Meldungen generiert wurden, so wird der Bezug auf die Ursprungsmeldung – wie schon in der Vertrauensstelle – gespeichert.

Gegenüber dem Schema der Vertrauensstelle werden die Pathologenmeldungen um die direkt vom Pathologen übertragenen tumorbezogenen Angaben ergänzt – temporär ist hierbei einer Pathologiemeldung evtl. keine Tumormeldung zugeordnet.

Im Gegensatz zur Vertrauensstelle benötigt die Registerstelle keine Informationen bezüglich der Standesämter bei einer Sterbemeldung. Aus diesem Grund existiert die Relation Standesamtinfos in der Registerstelle nicht.

Weiterhin ist die im konzeptionellen Schema nur transitiv über den Tumor definierte Beziehung zwischen Meldung und Patient jetzt explizit repräsentiert. Die Beziehung zwischen Sterbemeldung und Sterbefall wird aufgegeben, da sie sich implizit über den jeweiligen Patienten definiert.

Die c:c-Beziehung zwischen Tumor und Sterbefall bieten als Fremdschlüssel in der Relation Sterbefälle das (nach Best-Of-Ermittlung) validestes Grundleiden. Haben weitere Tumoren ursächlich zum Versterben beigetragen, so wird dies nicht als Fremdschlüssel, sondern über ein boolesches Attribut bei diesen vermerkt. Die amtliche Todesursachenstatistik muß exakt reproduziert werden können. Aus diesem Grund wird die amtliche Todesursache durch das Best-Of nicht verändert werden. Für die amtliche Todesursache existiert aus diesem Grund kein Tumor in Tumoren und daher referenziert ein Sterbefall eine Sterbemeldung.

Die Therapie besitzt jetzt eine direkte Zuordnung zum durch sie behandelten Tumor, die jeweils nach Durchführung des Tumormatch etabliert wird. Es erfolgt kein Abgleich zwischen unterschiedlichen gemeldeten Therapien zu einem Tumor, so daß evtl. auch widersprüchliche Angaben verschiedener Melder erhalten bleiben. Eine diesbezügliche Erweiterung des Konzeptes bleibt späteren Phasen des Registeraufbaus vorbehalten, zumal zunächst auch keine Routine-Auswertungen über die Therapie erfolgen werden.

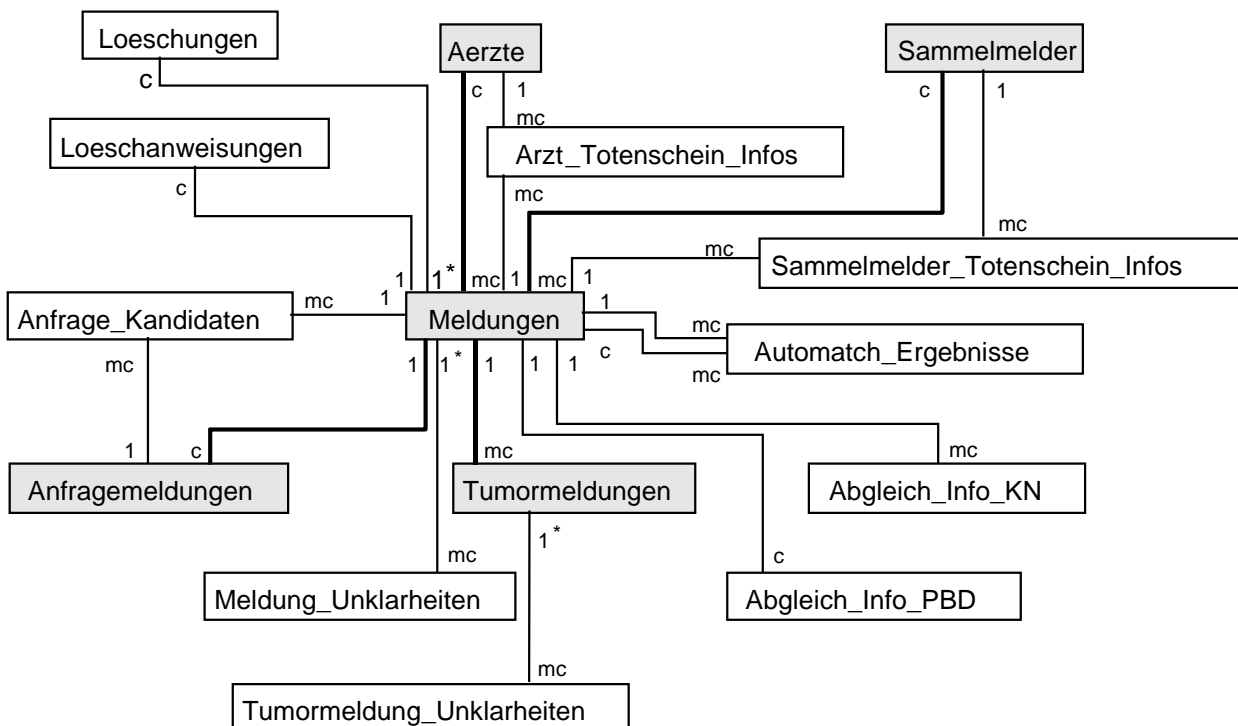


Abb. 2.3: Implementiertes ER-Schema der Registerstelle: Abgleich, Unklarheiten, Anfragen und Vitalstatusmeldung

Ähnlich werden Tätigkeiten, Familien- und Wohnortanamnesen behandelt, die jetzt – nach dem Patientenmatch – auch direkt den jeweiligen Patienten zugeordnet werden. Es erfolgt wiederum kein Abgleich zwischen unterschiedlichen Angaben von verschiedenen Meldern. Somit wird sich z.B. die angegebene Dauer der aktuellen Tätigkeit von Meldung zu Meldung natürlicherweise unterscheiden, so daß aus einem ausgeübten Beruf mehrere Einträge in der Relation Tätigkeiten resultieren. Diese Art der Verwaltung der aufgeführten Angaben ist derzeit noch ausreichend, da zunächst einmal die Validität der diesbezüglich gemeldeten Daten untersucht werden muß und auch noch nicht über diese ausgewertet wird. Durch Aufbewahrung der Originalangaben bleibt das Schema für spätere Umstrukturierungen offen.

Weiterhin werden Chiffretexte und Kontrollnummeraggregate zeilenweise in der Datenbank abgelegt, wodurch sich die 1:m- bzw. 1:mc-Beziehungen im Schema erklären.

Analog zur Vertrauensstelle (vgl. Abschnitt 1.1) werden auch in der Registerstelle temporär Tabellen zu Verwaltungszwecken, also für die Kommunikation mit der Vertrauensstelle sowie hier zusätzlich intern für den Datensatzabgleich, geführt (s. Abb. 2.3). Im Gegensatz zum Schema der Vertrauensstelle existieren hier auch Unklarheiten temporär ohne zugehörige Meldung bzw. Tumormeldung, so daß die referentielle Integrität nicht gewahrt ist. (Dies ist wiederum durch ein „*“ in Abb. 2.3 gekennzeichnet.) Löschanweisungen werden demgegenüber mit der betroffenen Meldung gelöscht. Schließlich werden in der Registerstelle nur Löschungen (nicht auch Auskunftsanforderungen) dauerhaft protokolliert (d.h. im wesentlichen lediglich gezählt), wobei der Bezug auf die Anfragemeldung keine Rolle spielt.

2.2 Relationen

Im folgenden werden die Relationen des Datenschemas in ähnlicher Form wie in Abschnitt 1.2 für die Vertrauensstelle genauer vorgestellt. Für Standardauswertungen in der Erprobungsphase relevante Attribute sind bei den personen- und tumorbezogenen Daten (Abschnitt 2.2.1 und 2.2.2) fett gedruckt. Hiervon abgesehen wird in diesem Bericht nicht weiter auf vorgesehene Auswertungen eingegangen. Die Ergebnisse dieser Auswertungen werden in dem Jahresbericht des Niedersächsischen Krebsregisters veröffentlicht, der erstmals 1998 für das Jahr 1997 erscheinen soll.

Anregungen zum Best-Of-Generator

In der Registerstelle werden durch Zusammenführung von Meldungen aus unterschiedlichen Quellen alle zu einem Patienten vorliegenden Informationen verschmolzen, wobei mittels eines sog. „Best-of-Generators“ [KSSM94] zu jedem Attribut die jeweils „beste“, d.h. verlässlichste Information aus den Einzelmeldungen extrahiert werden soll. Das Ziel der Definition von Best-of-Mechanismen besteht darin, entsprechende Regeln zu formulieren, die die folgenden zwei möglichen Konfliktarten berücksichtigen:

- Bei epidemiologischen Krebsregistern mit Meldungen aus unterschiedlichen Quellen kann davon ausgegangen werden, daß Informationen entsprechend der eigentlichen Aufgabe der meldenden Institution unterschiedlich hoch bewertet werden müssen. Medizinische Angaben, die z. B. von Pathologischen Instituten kommen, werden bzgl. ihrer Reliabilität wesentlich höher einzuschätzen sein als gleiche Informationen aus den Totenschei-

nen, wenn diese ohne Obduktion erhoben wurden. Solche Reliabilitätsgrade sollten jedoch nicht pauschal, sondern unter Berücksichtigung der einzelnen Attribute festgelegt werden, da eine Umkehrung des oben beschriebenen Beispiels bzgl. anderer erfaßter Daten (z. B. patientenbezogener Daten) durchaus vorstellbar ist.

- Ein weiterer Ansatz zur Festlegung von Regeln ist die Formulierung von Prioritäten auf den Ausprägungen von Attributen (sog. attributspezifische Best-of-Kriterien), die unabhängig vom Melder bestimmte Ausprägungen gegenüber anderen vorziehen. Ein einfaches Beispiel hierfür sind in einem Attribut über den Vitalstatus von Personen die Ausprägungen „unbekannt“, „nicht verstorben“, „verstorben“ (in aufsteigender Reihenfolge der Priorität).

Zusätzlich ist ein geschlossenes Verfahren zur Integration der beiden Regelwerke ein notwendiger Bestandteil für einen Best-of-Mechanismus, der in der Praxis dem Registerpersonal fundierte Entscheidungshilfen anbieten kann. Dieses muß jeweils festlegen, ob, wenn Melder unterschiedlichen Reliabilitätsgrades Ausprägungen mit unterschiedlicher Priorität liefern, erst der Reliabilitätsgrad oder erst die Priorität der Ausprägungen zu berücksichtigen ist.

Auf der Basis dieser Überlegungen werden in den Abschnitten 2.2.1 und 2.2.2 einige Ideen zur konkreten Definition eines Best-Of-Generators angeführt. Zusätzlich zur Beschreibung der Relationen wie in Abschnitt 1.2 werden zu jedem Attribut die jeweiligen, derzeit im Rahmen der Erprobungsphase als wesentliche Melder vorgesehenen meldenden Stellen (N = Nachsorgeleitstelle, P = Pathologe, T = Totenschein) angegeben. Die Reihenfolge der Melder gibt jeweils ihre Priorität beim Abgleich von Mehrfachmeldungen zum gleichen Patienten oder Tumor an (zunächst die Melder mit hoher Priorität). Ein „*“ kennzeichnet hier Attribute mit diesbezüglich speziellen Regelungen (die i.a. die Abhängigkeit vom Melder überwiegen und auch im Hinblick auf Mehrfachmeldungen des gleichen Melders zu sehen sind). Diese sind im Anschluß an die jeweilige Tabelle angeführt. Neben der Kompetenz des Melders bzgl. eines Attributs spielt hier die Erfassung des Status Quo zum ersten Diagnosezeitpunkt die zentrale Rolle, d.h. es soll keine zeitliche Fortschreibung der Angaben zur Erkrankung erfolgen. Sicherlich ist auch jede Angabe besser als ein NULL-Wert bzw. „fehlende Angabe / unbekannt“ oder ähnliches.

Generell hat der Best-Of-Generator nur einen Einfluß auf eine Default-Vorgabe, die das System dem Benutzer beim Vorliegen von Mehrfachmeldungen macht. Letztendlich ist die Wahl der Angaben zu einem tumor- oder patientenbezogenen Attribut aus den verschiedenen Meldungen dem Benutzer überlassen – insbesondere wenn sich etwa Priorität des Melders und attributspezifische Best-Of-Kriterien widersprechen. Weiterhin müssen die für den Tumor aggregierten Angaben zu den verschiedenen Attributen – ebenso wie die Meldungen selbst – untereinander konsistent bleiben, was (halb-)automatisch kontrolliert werden soll. Die in den folgenden Abschnitten aufgeführten Kriterien für das „Best-Of“ stellen zunächst einmal nur erste Anregungen dar und sollen im Laufe des nächsten Jahres noch verfeinert und konkretisiert werden.

Neben Angaben zum Best-Of-Generator wird im Anschluß an eine Tabelle jeweils ggf. zu einzelnen Attributen auf sich aus dem ER-Schema ergebende spezielle Aspekte der Datenmodellierung oder Konsistenzbedingungen eingegangen. Ausführlichere Erläuterungen zu den Attributen und ihren Ausprägungen finden sich in [Wie96].

2.2.1 Patienten

Relation Patienten

Attribut	Typ	Melder	Ausprägungen / Kommentare
<u>patient_id</u>	number(7)		Fortlaufende Nummer.
geschlecht	number(1)	NTP	- 1 (männlich) - 2 (weiblich) - 9 (fehlende Angabe / Sonstige)
geburtsjahr	number(4)	NTP	Vierstelliges Jahr – 0 für unbekannt.
geburtsmonat	number(2)	NT	Ein- bzw. zweistelliger Monat (1-12, 0 für unbekannt).
verstorben	number(1)	TN*	- 1 (Patient verstorben) - 2 (Patient nicht verstorben) - 9 (fehlende Angabe / unbekannt)
staatsangehoerigkeit	number(3)	NT*	Aktuelle (letzte) Staatsangehörigkeit sowie für Deutsche nach Möglichkeit Angabe des Bundeslandes des aktuellen Wohnortes. Kodierung nach der Klassifikation des Statist. Bundesamtes – 999 für unbekannt.
staat_version	number(4)	NT	Auflage der Landeskodierung (Jahr der Veröffentlichung, z.B. 1992) – 0 für unbekannt.
mehrling	number(1)	N	- 0 (kein Mehrling) - 1 (eineiiger Mehrling) - 2 (zweieiiger Mehrling) - 3 (Mehrling – unbekannt, ob ein- oder zweieiig) - 9 (fehlende Angabe / unbekannt)

Die in der Datenbank gespeicherten Patienten werden fortlaufend ab 1 durchnummeriert, d.h. der Primärschlüssel trägt keinerlei Semantik.

Liegt zu einem Patienten ein Totenschein vor, so wird stets auch das Flag *verstorben* gesetzt. Dies kann außerdem geschehen, wenn diesbezügliche Angaben schon auf einer Standardmeldung gemacht werden. Ein Best-Of zu diesem Attribut ist natürlich über die Ordnung „verstorben – nicht verstorben – unbekannt“ in fallender Priorität definiert.

Die *Staatsangehörigkeit* wird über die Zeit fortgeschrieben. Bei Änderungen wird also die letzte Angabe gespeichert.

Relation Sterbefaelle

Attribut	Typ	Melder	Ausprägungen / Kommentare
<u>patient_id</u>	number(7)		Schlüssel aus Patienten.
meldung_ref_amtliche_tu	number(8)	T	Fremdschlüssel aus Sterbetumormeldung. Definiert nur, wenn amtliche Todesursache (Grundleiden) auf dem Totenschein Tumor ist (sonst NULL).
tumormeldung_ref_amtliche_tu	number(1)	T	Fremdschlüssel aus Sterbetumormeldung. Definiert nur, wenn amtliche Todesursache (Grundleiden) auf dem Totenschein Tumor ist (sonst NULL).
tumor_ref_grundleiden	number(8)	T*	Fremdschlüssel aus Tumoren. Definiert nur, wenn das Grundleiden (Best-Of über alle Sterbemeldungen) ein Tumor ist (sonst NULL).
todesjahr	number(4)	T	Vierstelliges Jahr – 0 für unbekannt.
todesmonat	number(2)	T	Ein- bzw. zweistelliger Monat (1-12, 0 für unbekannt).

tod_tumorbedingt	number(1)	T	(Welcher Tumor zum Tod geführt hat, ist als Eintrag in der Relation Tumoren vermerkt.) - 1 (Tod tumorbedingt) - 2 (Tod nicht tumorbedingt) - 3 (fraglich, ob Tod tumorbedingt) - 9 (fehlende Angabe / unbekannt)
autopsie	number(1)	T*	- 1 (Autopsie durchgeführt) - 2 (Autopsie nicht durchgeführt) - 9 (fehlende Angabe / unbekannt)
wohnort	number(10)	T*	Wohnort zum Zeitpunkt des Todes: Gemeindekennziffer ¹⁹ bzw. entsprechende Kodierung auf höherer Aggregationsebene, falls nicht genauer möglich (vor allem für Pathologenmeldungen zunächst auf Landkreisebene) – 0 für unbekannt. ²⁰ Zukünftig Berücksichtigung regionaler Beobachtungseinheiten (vgl. Anhang), also z.B. Kodierung von Ortsteilen.
geocode_x	number(7)	T*	Gauß-Krüger-Koordinaten (vgl. Anhang; angestrebt sind 100 x 100 Meter, also zwei schließende Nullen) – 0 für unbekannt. Die x- (Rechts-) Koordinate beginnt mit dem Bezugsmeridian – zur Effizienzsteigerung soll evtl. alles auf den dritten Meridian (9° ö.L.) umgerechnet und in dieser Form abgelegt werden.
geocode_y	number(7)	T*	<s. geocode_x>

Für verstorbene Patienten wird aufgrund der Informationen aus entsprechenden Sterbemeldungen ein Eintrag in dieser Relation vorgenommen.

Evtl. kann aufgrund spezieller Erkenntnisse ein anderes *Grundleiden* als die amtliche Todesursache als die valideste Angabe zur Todesursache erscheinen. Ist dieses ein Tumor, wird an dieser Stelle eine Referenz auf Angaben zu diesem Tumor geboten. Zu der *amtliche Todesursache* ein Fremdschlüssel einer Sterbetumormeldung vorgehalten.

Der *Wohnort* zum Todeszeitpunkt wird (unabhängig vom Wohnort zum Diagnosezeitpunkt für Tumoren) als Best-of über alle Sterbemeldungen ermittelt.

Liegen widersprüchliche Angaben zur Durchführung einer *Autopsie* vor, so wird angenommen, daß eine Autopsie vorgenommen wurde.

2.2.2 Tumoren

Relation Tumoren

Attribut	Typ	Melder	Ausprägungen / Kommentare
<u>tumor_id</u>	number(8)		Fortlaufende Nummer.
patient_ref	number(7)		Fremdschlüssel aus Patienten.
wohnort	number(10)	NPT*	Wohnort zum Zeitpunkt der Diagnose: analog zur Relation Sterbefälle.

¹⁹ Bei der Nutzung des Datenschemas für Daten mit einem anderen Raumbezug (etwa im Fall von Stadtstaaten) sind – auf der Basis von Kodierungstabellen im Geo-Schema der Datenbank – zu den Gemeindekennziffern (vgl. Anhang) analoge Aggregationsebenen zu definieren.

²⁰ Wenn zumindest klar ist, daß der betreffende Patient aus dem Untersuchungsgebiet (also etwa Niedersachsen) stammt, so sollte dies auch so anstelle von 0 kodiert werden.

geocode_x	number(7)	NT*	<analog zu Sterbefälle>
geocode_y	number(7)	NT*	<analog zu Sterbefälle>
raucher_status	number(1)	N	Raucher-Status zum Zeitpunkt der Diagnose: - 1 (Nichtraucher) - 2 (Exraucher) - 3 (Raucher) - 9 (fehlende Angabe / unbekannt)
ex_raucher_jahr	number(4)	N	Das Jahr an dem der Raucher (s.o.) mit dem Rauchen aufgehört hat – 0 für unbekannt sowie für Raucher und Nichtraucher
diagnose_jahr	number(4)	NPT*	Vierstelliges Jahr – 0 für unbekannt.
diagnose_monat	number(2)	NPT*	Ein- bzw. zweistelliger Monat (1-12, 0 für unbekannt).
diagnose_anlass	number(1)	NT*	- 1 (Beschwerden) - 2 (Früherkennung) - 3 (Arbeitsmedizinische Untersuchung) - 4 (Nachsorge-Untersuchung) - 5 (Zufallsbefund) - 6 (Zufallsbefund bei Autopsie) - 7 (Sonstiges) - 9 (fehlende Angabe / unbekannt)
diagnose	char(6)	NPT*	Diagnose nach ICD mit 1 Nachkommastelle (z.B. „unbekannt“ nach ICD-9: 239.9 oder ICD-10: D48.9). Dreisteller i.a. als xxx.9 bzw. C/Dxx.9.
diagnose_version	number(2)	NPT	ICD-Version f. Diagnose (9, 10, ...).
histologie	number(4)	PNT*	Histologie nach ICD-O (z.B. „unbekannt“ nach ICD-O, 2. Auflage: 8000).
histologie_version	char(2)	PNT	ICD-O-Version f. Histologie, hierbei vorgestelltes D für die deutsche Ausgabe, also 1, 2, D1, D2.
lokalisation	char(6)	NPT*	Lokalisation nach dem Tumorlokalisationsschlüssel zur ICD-O mit mindestens 1 Nachkommastelle (z.B. „unbekannt“ nach ICD-O, 2. Auflage – entsprechend Version 4 des Lokalisationsschlüssels: C80.9).
lokalisation_version	number(2)	NPT	Version des Lokalisationsschlüssels (... , 3, 4, ...).
dignitaet	number(1)	PNT*	Malignitätsgrad aus dem ICD-O-Morphologie-Schlüssel (unabhängig von obigen Versionen): - 0 (gutartig) - 1 (unbest. Charakter / unsicher ob bös- oder gutartig) - 2 (in Situ) - 3 (bösaartig / Primärsitz) - 6 (bösaartig / Metastase) - 9 (fehl. Angabe / bösaartig, aber unbek. ob Metastase)
seite	number(1)	NPT	- 1 (rechts) - 2 (links) - 3 (beidseits) - 8 (trifft nicht zu) - 9 (fehlende Angabe / unbekannt)
diagnosesicherung	number(1)	PNT*	Beste Sicherung der Diagnose: - 1 (klinisch) - 2 (spezielle Diagnostik) - 3 (zytologisch) - 4 (histologisch) - 5 (autoptisch) - 6 (Sonstiges) - 9 (fehlende Angabe / unbekannt)

ausbreitung	number(1)	NPT*	<ul style="list-style-type: none"> - 0 (in Situ) - 1 (lokal begrenzt) - 2 (regionäre Lymphknoten / Nachbarschaft) - 3 (Fernmetastasen) - 4 (Systemerkrankung) - 9 (fehlende Angabe / unbekannt)
t	char(2)	PNT*	Tumorausbreitung. Ausprägungen: is, a, 0, 1, 2, 3, 4 und X für unbekannt.
ct	number(1)	PNT*	C-Faktor der Tumorausbreitung. Ähnlich der Diagnosesicherung: <ul style="list-style-type: none"> - 1 (klinisch) - 2 (spez. Diagnostik, evtl. auch Biopsie / Zytologie) - 3 (chirurg. Exploration mit Biopsie/Zytologie) - 4 (Tumorresektion mit patholog. Untersuchung) - 5 (autoptisch) - 9 (fehlende Angabe / unbekannt)
n	char(1)	PNT*	Ausbreitung in die Lymphknoten. Ausprägungen: 0, 1, 2, 3 und X für unbekannt.
m	char(1)	NPT*	Vorliegen von Fernmetastasen. Ausprägungen: 0, 1 und X für unbekannt.
tnm_version	number(2)	PNT	Version und Auflage des TNM-Schlüssels (etwa 41 für 4. Version, 1. Auflage oder 40 für 4. Version, unbekannte Auflage) für Angaben unter T, N – M sollte unabhängig von der Version sein – 0 für unbekannt.
grading	number(2)	PNT*	Differenzierungsgrad: <ul style="list-style-type: none"> - 1 (gut differenziert (G1)) - 2 (mäßig differenziert (G2)) - 3 (schlecht differenziert (G3)) - 4 (undifferenziert (G4)) - 5 (T-Zell-Lymphom) - 6 (B-Zell-Lymphom) - 7 (Null-Zell-Lymphom) - 9 (fehlende Angabe / n. bestimmbar / n. zutreffend) - 16 (Low grade (G1 oder G2)) - 17 (Medium grade (G2 oder G3)) - 18 (High grade (G3 oder G4)) - 19 (Grenzfall bzw. Borderline – nur bei Ovar)
tumorfolge	number(1)	NPT*	<ul style="list-style-type: none"> - 1 (erster Tumor) - 2 (zweiter Tumor) - 3 (weiterer Tumor) <p>Folgende Ausprägungen aus den Meldungen (vgl. Standardtumormeldungen) dürfen hier nicht vorkommen:</p> <ul style="list-style-type: none"> - 5 (früheres Tumorleiden vorhanden) - 6 (Metastase, Primärtumor unbekannt) - 9 (fehlende Angabe / unbekannt)
tod_tumorbedingt	number(1)	T*	<ul style="list-style-type: none"> - 1 (Tod durch diesen Tumor bedingt) - 2 (Tod nicht durch diesen Tumor bedingt) - 3 (fraglich, ob Tod durch diesen Tumor bedingt) - 9 (fehlende Angabe / unbekannt / nicht verstorben)
dco	number(1)	*	<ul style="list-style-type: none"> - 1 (DCO-Fall) - 2 (DCN-Fall) - 3 (sonst)
validitaet	number(1)	PNT*	Validität der Angaben zum Tumor: <ul style="list-style-type: none"> - 1 (valide / unauffällig / noch unbearbeitet) - 2 (auffällig, aber in Nachfrage bestätigt) - 3 (ungewöhnlich und (trotz Nachfrage) ungeklärt) - 4 (invalide / „unmöglich“ und ungeklärt)

Alle Tumoren werden – unabhängig vom jeweiligen Patienten – fortlaufend durchnummeriert. Der Primärschlüssel trägt also wiederum keinerlei Semantik.

Während der Erprobungsphase wird noch evaluiert werden müssen, ob es evtl. nötig sein könnte, durch ein zusätzliches Attribut mehrere (multiple) Tumoren eines Patienten als zur gleichen Erkrankung gehörig zu kennzeichnen (vgl. Tumormeldungen in Abschnitt 2.2.3). Derzeit entspricht ein Tumor genau einer Erkrankung (ein ICD-Kode) – bei der Festlegung des Best-Of muß ggf. eine Tumorausprägung (mit z.B. speziellem ICD-O-Kode) als „Haupttumor“ ausgewählt werden, falls die einzelnen Tumore nicht getrennt in die Fallzählung eingehen sollen.

Da sich die Angabe des *Wohnorts* auf den (frühesten) Diagnosezeitpunkt beziehen sollte, sind im Zweifelsfall Angaben aus früheren Meldungen beim Best-of vorzuziehen. *Gauß-Krüger-Koordinaten* werden zunächst einfach als Mittelpunkt des jeweiligen Wohnortes berechnet, da ihre Ermittlung aus der Patientenanschrift sowohl datenschutzrechtlich als auch technisch noch nicht geklärt ist.

Unabhängig vom jeweiligen Melder werden *Diagnosejahr* und *-monat* immer als der kleinste Diagnosezeitpunkt über alle Meldungen geführt. Aus der entsprechenden Meldung ergeben sich oft auch *Diagnoseanlaß*, *Ausbreitung*, *TNM* (unter Heranziehung der Klartexte) und *Grading*. Liegen die Diagnosezeitpunkte weniger als ein halbes Jahr²¹ auseinander, so sollten jedoch die oben angegebenen Prioritäten der Melder bzw. beim TNM-Kode ein evtl. vorhandenes p-Symbol das höhere Gewicht haben. Für Ausbreitung und TNM ist weiterhin ein fortgeschrittenes Stadium (innerhalb des halben Jahres) höher zu gewichten. Im Falle des *Grading* ist im Zweifelsfall gemäß [EWLP95] der höhere Grad zu wählen.

Die Festlegung der validesten *Diagnose* (und analog von *Histologie* und *Lokalisation*) basiert neben dem jeweiligen Melder auch auf Eigenheiten des ICD-Klassifikationsschemas, das die Kodierung auf verschiedenen Ebenen der Genauigkeit erlaubt (vgl. Anhang). In diese Relation sollte natürlich die genaueste Angabe aufgenommen werden, wobei auch die Diagnosesicherung einbezogen werden könnte. Weiterhin sollte nach Möglichkeit die Beschreibung des Tumors im „Anfangsstadium“, also zu Beginn der Behandlung, kodiert werden. Insgesamt bietet sich hier also die früheste histologische Untersuchung als Best-Of an. Für unspezifische Tumoren (entspricht „unbekannt“) sind spezielle Kodierungen im Klassifikationsschema vorgesehen. Wurde in der Meldung eine Metastase kodiert, erfolgt eine Umkodierung zu einem unbekanntem Primärtumor gemäß ICD – im ICD-O-Lokalisationsschlüssel wird weiterhin die Metastase kodiert (falls eine entsprechende Angabe vorhanden ist bzw. aus dem ICD-Kode gewonnen werden kann). Nach Möglichkeit sollen einheitlich ICD-10 und die 2. Auflage der ICD-O erfaßt bzw. eingegangene Meldungen auf Best-of-Ebene hierauf umkodiert werden.

Bei der *Dignität* wiegt die Angabe „bösartig“ oder sonst „in Situ“ stets schwerer als andere Ausprägungen. Metastasen sollten hier nicht auftreten.

Die Ausprägungen der *Diagnosesicherung* haben aufsteigende Priorität in der Reihenfolge „Sonstiges“, „klinisch“, „spez. Diagnostik“, „zytologisch“, „autoptisch“, „histologisch“. Im Falle der autoptischen Sicherung wird davon ausgegangen, daß eine histologische Untersuchung durchgeführt wurde. Angaben zu Diagnosesicherung und C-Faktor des TNM-Kodes (ggf. unter Berücksichtigung des p-Symbols des TNM) können wechselseitig zur Bildung des jeweiligen Best-of herangezogen werden. Hierbei entsprechen sich jeweils klinische und autoptische Sicherung,

²¹ Die Länge dieses Zeitraums ist sicher noch zu diskutieren.

eine spez. Diagnostik als Diagnosesicherung wird auf C2 abgebildet, zytologische Sicherung kann C2 oder C3 und histologische Sicherung C2, C3 oder C4 entsprechen. Umgekehrt wird aus C4 immer eine histologische Sicherung, aus C3 zytologisch oder histologisch und aus C2 spez. Diagnostik, zytologisch oder histologisch.²²

Die *Tumorfolge* ergibt sich im Best-of direkt aus der Reihenfolge der Tumoren in der Datenbank (gemäß Diagnosedatum). Somit sind nur die Ausprägungen „erster“, „zweiter“ und „weiterer Tumor“ zulässig. Tumoren mit gleichem Datum könnten in der Tumorfolge etwa nach medizinischer Relevanz oder anderen Aspekten geordnet werden (zunächst erfolgt bei fehlender medizinischer Beurteilung eine willkürliche Ordnung). Die meldungsspezifischen Angaben zum Vorliegen früherer Tumoren finden an dieser Stelle keine Berücksichtigung, sofern nicht aus zusätzlichen Klartexten neue Tumoren definiert werden (s.u.).

Alternative Möglichkeiten zur Definition der Tumorfolge – unter Einbeziehung meldungsspezifischer Angaben – wurden aufgrund unklarer bzw. realitätsferner Definitionen sowie der zu erwartenden geringen Validität der Angaben zur Tumorfolge verworfen (vgl. Abschnitt 1.2.1 unter Standardmeldungen). Zwei Varianten seien hier kurz erwähnt – sie basieren auf der Annahme, daß Inkonsistenzen in der Tumorfolge zwischen Datenbank und Meldungsangaben entweder durch das „Vergessen“ früherer Tumoren bei der Meldung oder – in der zweiten Variante – dadurch entstehen, daß späte Meldungen zu einem Tumor mit früherem ersten Diagnosezeitpunkt erst nach der Diagnose eines weiteren, späteren Tumors erfolgen. Eine gleichzeitige Berücksichtigung beider Möglichkeiten erlaubt keine eindeutige Definition.

In der ersten Variante definiert sich die Tumorfolge für jeden Tumor (beginnend mit dem gemäß Diagnosedatum ersten) als Maximum über Reihenfolge gemäß Diagnosedatum, Meldungsangaben auf allen zugehörigen Meldungen²³ und der bereits erfolgten Einordnung (erhöht um eine Stufe, falls nicht bereits „weiterer Tumor“) für alle früheren, also bereits betrachteten Tumoren.

Im anderen Fall wird mit folgenden Ausnahmen die Reihenfolge der Diagnosedaten übernommen: Liegt nur ein Tumor in der Datenbank vor, ist die Tumorfolge die höchste Numerierung aus den Meldungen.²⁴ Existieren zwei Tumoren in der Datenbank und ist bei einer Meldung zum ersten Tumor die Angabe "weiterer Tumor" kodiert, so wird der erste als zweiter und der zweite als weiterer Tumor kodiert. Analog wird der zweite Tumor zum weiteren Tumor hochgestuft, wenn dort in einer Meldung "weiterer Tumor" angegeben ist.

Klartexte zu früheren Tumoren, die in der Vertrauensstelle nicht zur Grundlage eigenständiger Meldungen gemacht wurden, können in das Best-of bereits bestehender Tumoren einfließen.

Die Angabe, ob der *Tumor den Tod bedingt* hat, muß konsistent sein zur Information in der Relation Sterbefaelle.

Die Ausprägung des Attributs *DCO* wird anhand der Menge der zum Tumor vorliegenden Meldungen und ihrer Eingangszeitpunkte in das Register festgelegt: Liegt nur ein Totenschein (Mel-demodus 5) vor, handelt es sich um einen DCO-Fall. Gibt es mehrere Meldungen zum Tumor,

²² Ohne weiteres Wissen sollte immer die jeweils schlechteste Sicherung angenommen werden.

²³ Metastasen und fehlende Angaben werden hierbei als Ersttumoren betrachtet, die Angabe "früheres Tumorleiden vorhanden" als Zweitumor.

²⁴ Behandlung von Metastasen, fehlenden Angaben sowie der Ausprägung „früheres Tumorleiden vorhanden“ wie oben.

von denen aber diejenige mit dem kleinsten Eingangsdatum wiederum der Totenschein ist, ist der Fall ein DCN-Fall.

Die *Validität* eines Tumors ist zu großen Teilen Ermessenssache der bearbeitenden Dokumentationskraft, die sich bei der Einstufung auf die Validität der betroffenen Tumormeldungen stützt (die höchste Validität wird i.a. übernommen).

2.2.3 Meldungen

Relation Meldungen

Attribut	Typ	Ausprägungen / Kommentare
meldung_id	number(8)	Fortlaufende Nummer.
meldung_ref	number(8)	Für Meldungen, die gemäß Klartextangaben zu früheren Tumoren aus einer anderen Meldung definiert wurden (vgl. Abschnitt 1.2.1), eine Referenz auf diese – sonst NULL.
patient_ref	number(7)	Fremdschlüssel aus Patienten – vor dem Abgleich NULL.
arzt_ref	number(5)	Fremdschlüssel aus Ärzte – ggf. NULL.
sammelmelder_ref	number(3)	Fremdschlüssel aus Sammelmelder – ggf. NULL.
vst_bearbeiter_ref	number(2)	Fremdschlüssel aus VST_Bearbeiter.
bearbeiter_ref	number(2)	Fremdschlüssel aus RST_Bearbeiter. Der Nachbearbeiter der Meldung – vor der Aufbereitung NULL; falls diese automatisch erfolgte, eine Referenz auf den Benutzer „Registerstelle“.
meldungsstatus	number(1)	- 1 (Neumeldung) - 2 (in Bearbeitung) - 4 (vorläufige Registermeldung) - 5 (Registermeldung) Folgende Ausprägung tritt in der Registerstelle nicht auf (vgl. Abschnitt 1.2.1): - 3 (in Nachfrage)
meldemodus	number(2)	- 1 (mit Information) - 2 (Ausnahmeregelung) - 3 (Pathologe) - 4 (Berufskrankheitenverfahren) - 5 (Todesbescheinigung) - 6 (Autopsiebefund) - 7 (klinischer Abschluß) - 10 (Widerspruchsmeldung) - 11 (Auskunftsanforderung Patient) - 12 (Auskunftsanforderung Forschung)
externe_referenz	char(15)	Externe Referenznummer, die nur für den Melder eine Semantik trägt (Identifikation des Patienten oder sogar der Meldung) – ggf. NULL.
eingangsdatum	date	Eingang in die Vertrauensstelle (Meldebogen oder Datenträger).
erfassungsdatum	date	Aufnahme in die Datenbank der Vertrauensstelle.
registrierungsdatum	date	Aufnahme in die Datenbank der Registerstelle.
matchdatum	date	Durchlauf von Automatch – vor dem Abgleich NULL.
aufbereitungsdatum	date	Letzte Nachbearbeitung der Abgleichergebnisse und Definition des Best-of – vorher NULL.

Die *Meldungs-ID* wird aus der Vertrauensstelle übernommen, es findet also keine Umnummerierung gegenüber den durch die Vertrauensstelle vergebenen Schlüsselnummern statt.

In Pathologenmeldungen werden aus Datenschutzgründen nach der Bearbeitung der Meldung der Verweis auf den *Arzt* und die zur Zusammenführung der Meldungsteile genutzte *externe Referenz* wieder gelöscht.

Über den *Meldungsstatus* werden aus der Vertrauensstelle transferierte neue Meldungen, in Bearbeitung befindliche (d.h. von Automatch abgeglichen, aber noch nicht aufbereitete), vorläufig sowie abschließend bearbeitete Meldungen unterschieden. Zu „vorläufigen Registermeldungen“ können Meldungen zu bisher noch nicht in der Datenbank gespeicherten Patienten werden, wenn sie eine geringe Vollständigkeit oder Validität aufweisen. Es wird jedoch von diesen Meldungen noch kein Best-of gebildet. Ggf. können sie noch bis zum Ablauf der 3-Monatsfrist in der Vertrauensstelle nachgefragt werden (etwa, wenn eine weitere Meldung zum gleichen Patienten eintrifft). Nach Ablauf einer noch genauer zu definierenden Frist erfolgen automatisch Erzeugung des Best-of und Umwandlung in eine Registermeldung, oder die Meldung wird gelöscht (z.B. wenn sie zu lückenhafte Informationen aufweist).

Widerspruchs- und Auskunftsmeldungen werden nach Bearbeitung wieder aus der Datenbank entfernt. Ebenso Totenscheine, die weder einem bestehenden Patienten zugeordnet werden konnten noch Angaben zu Tumoren enthalten. Evtl. könnte später für Auswertungen eine Sicht als Einschränkung auf Registermeldungen definiert werden.

Der *Meldemodus* lehnt sich an das im derzeitigen Entwurf zum Niedersächsischen Krebsregistergesetz festgelegte Meldemodell an. 1 oder 2 steht für Standardmeldungen; 5, 6 oder 7 für Sterbemeldungen sowie 3 oder 4 für (evtl. um eine Berufsanamnese erweiterte) Pathologenmeldungen. Hierbei sind 4 und 6 zunächst noch nicht relevant; die Meldemodi 10 bis 12 dienen für Anfragemeldungen.

Die Protokollierung des *Bearbeiters* einer Meldung ist – ebenso wie die des Bearbeiters bzw. Erfassers in der Vertrauensstelle – datenschutzrechtlich noch zu diskutieren (siehe Seite 7). Der Bearbeiter repräsentiert – analog zur Vertrauensstelle – die letzte Beschäftigung (inkl. Nachfragen) mit der Meldung.

Relation Tumormeldungen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>tumormeldung_id</u>	number(1)	Fortlaufende Numerierung der Tumoren einer Meldung.
multipler_tumor_id	number(1)	Fortlaufende Numerierung der „unabhängigen“ Tumoren einer Meldung.
tumor_ref	number(8)	Fremdschlüssel aus Tumoren – vor dem Abgleich NULL.
validitaet	number(1)	<s. Tumoren>

Die (*Tumor-*)*Meldungs-IDs* werden aus der Vertrauensstelle übernommen.

Relation Anfragemeldungen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.

geschlecht	number(1)	<s. Patienten>
geburtsjahr	number(4)	<s. Patienten>
geburtsmonat	number(2)	<s. Patienten>
wohntort	number(10)	<s. Tumoren>

Relation Standardmeldungen

Attribut	Typ	Ausprägungen / Kommentare
meldung_id	number(8)	Schlüssel aus Meldungen.
geschlecht	number(1)	<s. Patienten>
geburtsjahr	number(4)	<s. Patienten>
geburtsmonat	number(2)	<s. Patienten>
verstorben	number(1)	<s. Patienten>
staatsangehoerigkeit	number(3)	<s. Patienten>
staat_version	number(4)	<s. Patienten>
mehrling	number(1)	<s. Patienten>
wohntort	number(10)	<s. Tumoren>
geocode_x	number(7)	<s. Tumoren>
geocode_y	number(7)	<s. Tumoren>
raucher_status	number(1)	<s. Tumoren>
ex_raucher_jahr	number(4)	<s. Tumoren>
diagnose_jahr	number(4)	<s. Tumoren>
diagnose_monat	number(2)	<s. Tumoren>
patienten_bemerkung	number(1)	Hat der Patient zu seiner Krebserkrankung weitere Anmerkungen gemacht: - 1 (ja) - 2 (nein) - 9 (nicht befragt / fehlende Angabe / unbekannt)

In der Relation Wohnort_Texte findet sich ggf. ein Klartext zum Wohnort, und in der Relation Fruehere_Tumoren_Texte sind ggf. Spezifikationen früherer Tumoren abgelegt.

Relation Standardtumormeldungen

Attribut	Typ	Ausprägungen / Kommentare
meldung_id	number(8)	Schlüssel aus (Tumor-)Meldungen.
tumormeldung_id	number(1)	Schlüssel aus Tumormeldungen.
diagnose_anlass	number(1)	<s. Tumoren>
diagnose	char(6)	<s. Tumoren>
diagnose_version	number(2)	<s. Tumoren>
histologie	number(4)	<s. Tumoren>
histologie_version	char(2)	<s. Tumoren>
lokalisierung	char(6)	<s. Tumoren>
lokalisierung_version	number(2)	<s. Tumoren>
dignitaet	number(1)	<s. Tumoren>
seite	number(1)	<s. Tumoren>
diagnosesicherung	number(1)	<s. Tumoren>
ausbreitung	number(1)	<s. Tumoren>
grading	number(2)	<s. Tumoren>
tumorfolge	number(1)	- 1 (erster Tumor) - 2 (zweiter Tumor) - 3 (weiterer Tumor) - 5 (früheres Tumorleiden vorhanden) - 6 (Metastase, Primärtumor unbekannt) - 9 (fehlende Angabe / unbekannt)

In der Relation Diagnose_Texte findet sich ein Klartext zur Diagnose, und in der Relation TNM_Texte sind ggf. Angaben zum TNM-Kode.

Relation Pathologenmeldungen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
geschlecht	number(1)	<s. Patienten>
geburtsjahr	number(4)	<s. Patienten>
wohntort	number(10)	<s. Tumoren>, vorerst Kodierung des Landkreises, später einmal (für die Dauerphase) regionale Beobachtungseinheit.
diagnose_jahr	number(4)	<s. Tumoren>, Zeitpunkt der histolog. Erfassung.
diagnose_monat	number(2)	<s. Tumoren>

Relation Pathologentumormeldungen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus (Tumor-)Meldungen.
<u>tumormeldung_id</u>	number(1)	Schlüssel aus Tumormeldungen.
diagnose_jahr	number(4)	<s. Tumoren>
diagnose_monat	number(2)	<s. Tumoren>
diagnose	char(6)	<s. Tumoren>
diagnose_version	number(2)	<s. Tumoren>
histologie	number(4)	<s. Tumoren>
histologie_version	char(2)	<s. Tumoren>
lokalisation	char(6)	<s. Tumoren>
lokalisation_version	number(2)	<s. Tumoren>
dignitaet	number(1)	<s. Tumoren>
seite	number(1)	<s. Tumoren>
diagnosesicherung	number(1)	<s. Tumoren>
ausbreitung	number(1)	<s. Tumoren>
grading	number(2)	<s. Tumoren>
tumorfolge	number(1)	<s. Standardtumormeldungen>

In der Relation Diagnose_Texte findet sich ein Klartext zur Diagnose, und in der Relation TNM_Texte sind ggf. Angaben zum TNM-Kode.

Die Zusammenführung von Pathologen- und Pathologentumormeldungen gemäß peripherer Teilanonymisierung erfolgt, indem nach dem Transfer des Teils unter Pathologenmeldungen (aus der Vertrauensstelle) der direkt vom Pathologen gemeldete Teil unter Pathologentumormeldungen über die externe Referenz (s. Relation Meldungen) zugeordnet und dazugespielt wird. Die Datumsangaben unter Meldungen beziehen sich also auf den von der Vertrauensstelle transferierten Meldungsteil.

Relation Sterbemeldungen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
geschlecht	number(1)	<s. Patienten>
geburtsjahr	number(4)	<s. Patienten>

geburtsmonat	number(2)	<s. Patienten>
staatsangehoerigkeit	number(3)	<s. Patienten>
staat_version	number(4)	<s. Patienten>
wohnort	number(10)	<s. Tumoren>
geocode_x	number(7)	<s. Tumoren>
geocode_y	number(7)	<s. Tumoren>
todesjahr	number(4)	<s. Sterbefälle>
todesmonat	number(2)	<s. Sterbefälle>
unmittelbare_tu_icd	char(6)	Kodierung der unmittelbaren Todesursache nach ICD – ICD-9: 799.9 für „unbekannt“, nach ICD-10: R99.9.
unmittelbare_tu_version	number(2)	ICD-Version zu unmittelbarer Todesursache.
grundleiden_icd	char(6)	Analog amtliche Todesursache nach ICD.
grundleiden_version	number(2)	Analog ICD-Version zu Grundleiden.
autopsie	number(1)	<s. Sterbefälle>

In der Relation Wohnort_Texte findet sich ggf. ein Klartext zum Wohnort. Außerdem stehen eventuelle Klartexte zur unmittelbaren Todesursache sowie zum Grundleiden in den entsprechenden Klartext-Relationen. Beim Grundleiden findet sich ggf. eine Beschreibung der Epikrise, also der vollständigen Kausalkette der Todesursachen, falls dies erforderlich sein sollte. In der Relation Fruehere_Tumoren_Texte werden ggf. Spezifikationen früherer Tumoren abgelegt.

Es werden nur Totenscheine zu Patienten, die bereits in der Datenbank vorliegen, sowie solche mit Krebs in der Todesursachenkette in die Datenbank aufgenommen.

Zur Kodierung von Autopsiebefunden s. Abschnitt 1.2.1 unter Sterbemeldungen.

Relation Sterbetumormeldungen

Attribut	Typ	Ausprägungen / Kommentar
meldung_id	number(8)	Schlüssel aus (Tumor-)Meldungen.
tumormeldung_id	number(1)	Schlüssel aus Tumormeldungen.
art_der_todesursache	number(1)	- 1 (unmittelbare Todesursache) - 2 (als Folge von) - 3 (Grundleiden) - 4 (weitere Erkrankung) - 5 (Epikrise) - 9 (fehlende Angabe / unbekannt)
diagnose_jahr	number(4)	<s. Tumoren>, bei Tod angegebener Erkrankungszeitpunkt.
diagnose_monat	number(2)	<s. Tumoren>
diagnose_anlass	number(1)	<s. Tumoren>
diagnose	char(6)	<s. Tumoren>
diagnose_version	number(2)	<s. Tumoren>
histologie	number(4)	<s. Tumoren>
histologie_version	char(2)	<s. Tumoren>
lokalisierung	char(6)	<s. Tumoren>
lokalisierung_version	number(2)	<s. Tumoren>
dignitaet	number(1)	<s. Tumoren>
seite	number(1)	<s. Tumoren>
diagnosesicherung	number(1)	<s. Tumoren>
ausbreitung	number(1)	<s. Tumoren>
grading	number(2)	<s. Tumoren>
tumorfolge	number(1)	<s. Standardtumormeldungen>
tod_tumorbedingt	number(1)	<s. Tumoren>

In der Relation *Diagnose_Texte* findet sich ein Klartext zur Diagnose, und in der Relation *TNM_Texte* sind ggf. Angaben zum TNM-Kode.

In dem *Diagnosedatum* in der Sterbetumormeldung ist der Zeitpunkt der Erkrankung, wenn dieser dem Melder zum Zeitpunkt des Todes bekannt ist. Beispielsweise kann auf einem Totenschein vermerkt werden, wann der Patient an dem Tumor erkrankt ist. Der DCO-Status eines Tumors bleibt durch diese Vorgehensweise unangetastet.

2.2.4 Weitere meldungsbezogene Daten

Relation Taetigkeiten

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>taetigkeit_id</u>	number(1)	Fortlaufende Nummer der Tätigkeiten einer Meldung.
patient_ref	number(7)	Fremdschlüssel aus Patienten – vor Abgleich NULL.
beruf	number(4)	Kodierung der Tätigkeit nach der Klassifikation der Berufe der Bundesanstalt für Arbeit – 9911 für unbekannt.
beruf_version	number(4)	Version der Kodierung (Jahr des Erscheinens der Auflage) – 0 für unbekannt.
taetigkeit_dauer	number(4)	Dauer der Berufsausübung (bis zum Zeitpunkt der Meldung) in Monaten – 0 für unbekannt.
letzte_taetigkeit	number(1)	War diese Tätigkeit die letzte vor der Diagnose des jeweiligen Tumors der zugehörigen Meldung? - 1 (ja) - 2 (nein) - 9 (fehlende Angabe / unbekannt)
laengste_taetigkeit	number(1)	War diese Tätigkeit die längste vor der Diagnose des jeweiligen Tumors der zugehörigen Meldung? - 1 (ja) - 2 (nein) - 9 (fehlende Angabe / unbekannt)

In der Relation *Beruf_Texte* findet sich ggf. ein Klartext zum Beruf. Wahrscheinlich sollen Klartexte jedoch i.a. nicht erfaßt werden.

Die *Tätigkeits-ID* wird aus der Vertrauensstelle übernommen.

Die Informationen zu Tätigkeiten werden aus jeder Meldung unverändert übernommen, es erfolgt keine Zusammenführung der Angaben zu einem Patienten und auch keine Fortschreibung etwa der Attribute *Tätigkeitsdauer* oder *längste_Tätigkeit*. Über die Definition des Fremdschlüssels zum jeweiligen Patienten könnte modelliert werden, ob eine Tätigkeit als für Auswertungen relevant angesehen wird. (Dieser könnte z.B. nur für Angaben zum ersten Tumor eines Patienten ungleich NULL sein. Hiervon wird jedoch zunächst noch kein Gebrauch gemacht, d.h. der Fremdschlüssel wird immer definiert.) Der Bezug zum jeweiligen Tumor wird nicht explizit repräsentiert. Hier besteht noch das Bedürfnis, Erfahrungen zu sammeln und aufgrund dieser Erfahrungen im Verlauf der weiteren Registeraufbaus ggf. Modifikationen oder Erweiterungen des Datenschemas vorzunehmen.

Relation Wohnortanamnesen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
patient_ref	number(7)	Fremdschlüssel aus Patienten – vor Abgleich NULL.
wohnung_seit_min	number(4)	Angabe eines Bereichs von Jahreszahlen (Unterstützung unscharfer Angaben, i.a. Min = Max), seit dem der aktuelle Wohnort Gültigkeit hat – 0 für unbekannt.
wohnung_seit_max	number(4)	<s. wohnort_seit_min>
geburtsort_gkz	number(10)	Gemeindekennziffer des Geburtsortes (auch außerhalb Niedersachsens, falls vorhanden) – 0 falls undefiniert (außerhalb Deutschlands) oder unbekannt.
geburtsland	number(3)	Ländercode des Statistischen Bundesamtes für das Geburtsland (möglichst inklusive Kodierung des Bundeslandes in Deutschland) – 999 für unbekannt.
geburtsort_text	char(80)	Klartext für Geburtsort und -land.
aufwuchs_gkz	number(10)	Analog zum Geburtsort Gemeindekennziffer des Ortes des Aufwachsens – 0 falls undefiniert oder unbekannt.
aufwuchs_land	number(3)	Analog zum Geburtsland Ländercode des Statistischen Bundesamtes für den Ort des Aufwachsens – 999 für unbekannt.
aufwuchs_text	char(80)	Klartext für Ort und Land des Aufwachsens.
laengster_aufenthalt_ort	number(10)	Kodierung des Ortes des längsten Aufenthalts analog zum Wohnort in Relation Tumoren.
laengster_aufenthalt_land	number(3)	Analog zum Geburtsland Ländercode des Statistischen Bundesamtes für das Land des bisher überwiegenden Aufenthalts – 999 für unbekannt.
laengster_aufenthalt_text	char(80)	Klartext für Ort und Land des bisher überwiegenden Aufenthalts (ohne Anschrift).
staat_version	number(4)	Jeweils verwendete Version der Länderkodierungen – 0 für unbekannt.

Wiederum erfolgt keine Zusammenführung der Angaben aus verschiedenen Meldungen.

Relation Familienanamnesen

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>familie_id</u>	number(1)	Fortlaufende Nummer der Tumoren in der Familie auf einer Meldung.
patient_ref	number(7)	Fremdschlüssel aus Patienten – vor Abgleich NULL.
verwandtschaft	number(1)	Grad der Verwandtschaft: - 0 (Enkel) - 1 (Kinder) - 2 (Geschwister) - 3 (Eltern) - 4 (Neffen) - 5 (Großeltern) - 6 (Onkel / Tanten) - 7 (Cousins / Cousinen) - 8 (Sonstige) - 9 (fehlende Angabe / unbekannt)
geschlecht	number(1)	<s. Patienten>
diagnose	char(6)	Diagnose der Krebserkrankung nach ICD.
diagnose_version	number(2)	<s. Tumoren>

Die *Familien-ID* wird aus der Vertrauensstelle übernommen.

Auch hier wird kein Best-of erzeugt.

Relation Therapien

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus (Tumor-)Meldungen.
<u>tumormeldung_id</u>	number(1)	Schlüssel aus Tumormeldungen.
<u>therapie_id</u>	number(1)	Fortlaufende Nummer der Therapien zu einem Tumor auf einer Meldung.
tumor_ref	number(8)	Fremdschlüssel aus Tumoren – vor Abgleich NULL.
therapieart	number(1)	- 1 (Operation) - 2 (Radiatio) - 3 (Chemotherapie) - 4 (Hormontherapie) - 5 (Immuntherapie) - 6 (Knochenmarkstransplantation) - 7 (Sonstige) - 9 (fehlende Angabe / unbekannt)
therapieziel	number(1)	- 1 (kurativ) - 2 (palliativ) - 3 (adjuvant) - 4 (supportiv) - 5 (neoadjuvant) - 6 (explorativ) - 7 (Sonstige) - 9 (fehlende Angabe / unbekannt)
therapiestatus	number(1)	- 1 (durchgeführt) - 2 (nicht durchgeführt) - 3 (vorgesehen) - 4 (verweigert)

Die *Therapie-ID* wird aus der Vertrauensstelle übernommen.

Beim *Therapiestatus* ist keine fehlende Angabe vorgesehen – entsprechende Einträge werden gar nicht erst gemacht.

Da wiederum keine Zusammenführung der verschiedenen Angaben zur Therapierung eines Tumors erfolgt, spielt der Best-Of-Generator auch hier keine Rolle.

2.2.5 Texte und Chiffrate

Klartexttabellen zu den Meldungen (**Wohnort_Texte**, **Todesursache_Texte**, **Grundleiden_Texte**, **Beruf_Texte**, **Fruehere_Tumoren_Texte**, **Diagnose_Texte**, **TNM_Texte**) sind exakt wie in der Vertrauensstelle (s. Abschnitt 1.2.3) definiert. Klartexte zu Diagnosen und TNM-Kodierung werden in der Registerstelle zusätzlich den Pathologentumormeldungen zugeordnet.

Wie schon in der Vertrauensstelle, können auch in der Registerstelle bei der Meldungsbearbeitung Kommentare – hier nun zu Meldungen, Tumoren und Patienten – gespeichert werden. Die

Relationen **Kommentar_Meldung** und **Kommentar_Tumormeldung** werden aus der Vertrauensstelle übernommen.

Relation **Kommentar_Tumor**

Attribut	Typ	Ausprägungen / Kommentar
<u>tumor_id</u>	number(8)	Schlüssel aus Tumoren.
<u>zeile</u>	number(3)	Blöcke eines Klartextes ab 1.
text	char(80)	Der Text.

Relation **Kommentar_Patient**

Attribut	Typ	Ausprägungen / Kommentar
<u>patient_id</u>	number(7)	Schlüssel aus Patienten.
<u>zeile</u>	number(3)	Blöcke eines Klartextes ab 1.
text	char(80)	Der Text.

Chiffretext (wie schon in der Vertrauensstelle – s. Abschnitt 1.2.4) und Kontrollnummeraggregat einer Meldung werden gleichermaßen in Zeilen mit jeweils (abgesehen von der letzten Zeile) 80 Zeichen aufgeteilt, die getrennt in einzelnen Tupeln zweier Relationen gespeichert werden:

Relation **Chiffretexte**

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>zeile</u>	number(3)	Fortlaufende Numerierung der Textzeilen ab 1.
schlüsseltext	char(80)	Schlüsseltext.

Relation **Kontrollnummeraggregate**

Attribut	Typ	Ausprägungen / Kommentare
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>zeile</u>	number(3)	Fortlaufende Numerierung der Textzeilen ab 1.
schlüsseltext	char(80)	Schlüsseltext.

2.2.6 Kodierungstabellen für Attribute, Dokumentationskräfte und Melder

Die verschiedenen **Kodierungstabellen** zu allen Attributen vom Aufzählungstyp (inkl. **K_ICD**, **K_Histologie**, **K_Lokalisation**, **K_Beruf**, **K_Staat** und die jeweiligen **Versionen**) sind exakt wie in der Vertrauensstelle (s. Abschnitt 1.2.5) definiert.

Wohnorte mit den zugehörigen Gemeindecennziffern werden nicht in einer speziellen Kodierungstabelle, sondern im Schema der Geodaten (Relation **Gebiete**) verwaltet (s. Abschnitt 4.2.2).

VST- und **RST_Bearbeiter** werden aus der Vertrauensstelle übernommen.

Die Relationen zu den Meldern (**Ärzte**, **Sammelmelder** und **Institutionen**) sind lediglich um die Bankverbindung gekürzt.²⁵ Es erfolgt je nach Bedarf – getrennt von übermittelten Meldungen²⁶ – ein Update aus den entsprechenden Relationen der Vertrauensstelle. Da ein direkter Kontakt der Registerstelle mit dem meldenden Arzt (etwa zur Rückfrage) nicht vorgesehen ist, sondern die Kommunikation nach dem derzeitigen Stand der Überlegungen stets über die Vertrauensstelle laufen soll, ist die Relevanz der Attribute *Telefon*, *Aktenzeichen* und *KV-Nummer* in der Registerstelle noch fraglich.

Eine Ausnahme bilden wahrscheinlich Pathologenmeldungen, da hier datenschutzrechtliche Bedenken gegen eine dauerhafte Speicherung des Melders bestehen.

2.2.7 Hilfsrelationen

Zur effizienteren Auswertung der Daten werden Ausschnitte der Datenbasis automatisiert redundant in Hilfsrelationen gehalten, die speziell auf bestimmte Anfragegruppen zugeschnitten sind.

Relation Erweiterte_Tumordaten

Attribut	Typ	Ausprägungen / Kommentare
tumor_id	number(8)	Schlüssel aus Tumoren.
patient_ref	number(7)	Fremdschlüssel aus Patienten.
geschlecht	number(1)	<s. Patienten>
wohntort	number(10)	Fremdschlüssel aus Gebiete: Interne Gebietsnummer des Wohnortes aus der Relation Tumoren. ²⁷
geocode_x	number(7)	<s. Tumoren>
geocode_y	number(7)	<s. Tumoren>
diagnose_jahr	number(4)	<s. Tumoren>
diagnose_monat	number(2)	<s. Tumoren>
diagnose_zeitpunkt	number(5)	12 * diagnose_jahr + diagnose_monat. Falls Monat unbekannt: – (12 * diagnose_jahr + 6). Falls Jahr und Monat unbekannt, dann 0. Temporär beim Laden auch NULL.
diagnose_alter	number(3)	Aus Geburts- und Diagnosejahr sowie -monat berechnet. Liegt zu einem der beiden Daten der Monat nicht vor, werden nur die Jahre zur Altersberechnung herangezogen. -1 für unbekannt.
diagnose_anlass	number(1)	<s. Tumoren>
diagnose	char(6)	<s. Tumoren>, Diagnose vierstellig nach ICD-10.
histologie	number(4)	<s. Tumoren>, Histologie nach ICD-O, 2. (englische) Auflage.
histologie_d	number(5)	10 * histologie + dignitaet. Temporär beim Laden auch NULL.

²⁵ Es wird derzeit auch die Möglichkeit diskutiert, die Abrechnung mit den Meldern durch die Registerstelle vornehmen zu lassen. In diesem Fall ist hier natürlich eine Anpassung erforderlich.

²⁶ Natürlich muß die Referenz auf den Melder für jede Meldung auch in der Registerstelle definiert sein.

²⁷ Aus Effizienzgründen wurde in dieser Relation auf die Definition der Fremdschlüssel auf die Relation Gebiete verzichtet, da beim Laden der Tabelle hier temporär direkt die Wohnorte aus der Relation Tumoren gespeichert werden.

dignitaet	number(1)	<s. Tumoren>
diagnosesicherung	number(1)	<s. Tumoren>
ausbreitung	number(1)	<s. Tumoren>
tumorfolge	number(1)	<s. Tumoren>
sterbe_jahr	number(4)	<s. Sterbefälle>, ggf. NULL.
sterbe_monat	number(2)	<s. Sterbefälle>, ggf. NULL.
sterbe_zeitpunkt	number(5)	Analog zu Diagnosezeitpunkt, ggf. NULL.
sterbe_alter	number(3)	Aus Geburts- und Todesdatum berechnet (s. Diagnosealter), ggf. NULL.
dco	number(1)	<s. Tumoren>
validitaet	number(1)	<s. Tumoren>
sterbe_wohnort	number(10)	Fremdschlüssel aus Gebiete: Interne Gebietsnummer des Wohnortes aus der Relation Sterbefälle (vgl. Wohnort). Genau dann belegt, wenn die jeweilige Person verstorben ist, sonst NULL.
sterbe_geocode_x	number(7)	<s. Sterbefälle>, ggf. NULL.
sterbe_geocode_y	number(7)	<s. Sterbefälle>, ggf. NULL.
tod_tumorbedingt	number(1)	<s. Tumoren>. Genau dann, wenn dieses Feld mit „ja“ belegt ist, wird der Fall als Mortalitätsfall zum durch das Tupel repräsentierten Tumor mitgezählt.
tod_tumorbedingt_amtlich	number(1)	<s. Tumoren>
t	char(2)	<s. Tumoren>
n	char(1)	<s. Tumoren>
m	char (1)	<s. Tumoren>

Diese Relation wird zum effizienten Zugriff auf die Tumordaten benötigt und automatisch aus den Relationen Tumoren, Patienten und Sterbefälle ermittelt.

Die *Gauß-Krüger-Koordinaten* werden hier auch schon einmal abgelegt, obwohl noch keine konkreten Auswertungen damit geplant sind.

Die Angabe, ob der *Tod tumorbedingt* ist, entscheidet darüber, ob der jeweilige Tumor mit in die Mortalität eingeht. Das Feld *Tod tumorbedingt* gibt an, ob der Tod nach dem Abgleich in der Registerstelle als tumorbedingt angenommen werden kann. Die amtliche Todesursache, aus der die amtliche Todesursachenstatistik gebildet wird, kann von dem durch das best of ermittelte Grundleiden abweichen. Das Feld *Tod tumorbedingt amtlich* kennzeichnet die Mortalitätsfälle in denen die amtliche Todesursache vermerkt ist. Im allgemeinen sollten zu einem Tumor, der zum Tode des Patienten geführt hat, zwei Tupel in der Relation *Erweiterte_Tumordaten* stehen. Ein Tupel mit den Inzidenz- und Mortalitätsdaten gemäß des best of und ein Tupel allein mit den amtlichen Angaben zur Todesursache.

2.2.8 Relationen für Datensatzabgleich und Verwaltung

Während des Datensatzabgleichs werden temporär Relationen zu den abgeglichenen Datensätzen angelegt. Diese Relationen enthalten die Kontrollnummern und weitere personenidentifizierende Angaben der betrachteten Meldungen sowie das Ergebnis des Matchlaufs. Der genaue Ablauf des Abgleichs wird hier nicht beschrieben – hierzu ist der gesonderter Bericht [HDW97] erschienen.

Relation Abgleich_Info_KN

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>kn_id</u>	number(2)	Index der Kontrollnummer (1-22).
wert	char(23)	Kontrollnummerntext.

In dieser Relation werden temporär aus den Kontrollnummeraggregaten extrahierte Kontrollnummern zu Neu- und Registermeldungen abgelegt.

Relation Abgleich_Info_PBD

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
<u>kandidat_id</u>	number(8)	Interne Hilfsnumerierung zum leichteren Selektieren.
geschlecht	number(1)	<s. Patienten>
geburtsjahr	number(4)	<s. Patienten>
geburtsmonat	number(2)	<s. Patienten>
gemeindekennziffer	number(10)	<s. Tumoren>. Hier sind Überlegungen anzustellen, ob bzw. in welcher Form der Abgleich nach Einführung regionaler Beobachtungseinheiten auch unter deren Berücksichtigung erfolgen soll.
treffer	number(1)	- 0 (Meldung kein Treffer) - 1 (Meldung ist Treffer)

Hier werden weitere für den Abgleich erforderliche Attribute (personenbezogene Daten) zu Neu- und Registermeldungen temporär abgelegt.

Relation Automatch_Ergebnisse

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen – abgegliche Meldung.
<u>ergebnis_id</u>	number(2)	Fortlaufende Nummer der zugeordneten Meldung.
position	number(2)	Position der zugeordneten Meldung (in der Reihenfolge der von AutoMatch vergebenen Gewichte – der beste Match hat die Position 1) bzw. 0, falls keine korrespondierende Meldung von AutoMatch gefunden wurde.
meldung_ref	number(8)	Fremdschlüssel aus Meldungen – Identifikation der zugeordneten Meldung bzw. NULL bei Mißerfolg.
gewicht	number(5,3)	Abgleichgewicht bzw. NULL bei Mißerfolg.
blockvariable	number(1)	Gibt an, über welche Blockvariable gematched wurde, bzw. NULL bei Mißerfolg.
bearbeitungsstatus	number(1)	- 0 (normal / nicht in Bearbeitung) - 1 (in Bearbeitung)

Diese Relation enthält die Ergebnisse eines Automatch-Durchlaufs.

Der *Bearbeitungsstatus* wird nur von Automatch intern bei der Hüllenbildung verwendet.

Relation Automatch_Huellen

Attribut	Typ	Ausprägungen / Kommentar
<u>huelle_id</u>	number(8)	Fortlaufende Nummer.
bearbeitungsstatus	number(1)	- 0 (normal / nicht in Bearbeitung) - 1 (in Bearbeitung)

Hier dokumentiert der *Bearbeitungsstatus*, ob sich die betreffende Meldung gerade in der Nachbearbeitung befindet – somit wird die Mehrbenutzerfähigkeit des Systems unterstützt.

Relation Automatch_Huellen_Elemente

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel aus Meldungen.
huelle_ref	number(8)	Fremdschlüssel aus Automatch_Huellen.
kn_pseudoauspraegungen	char(22)	Ein String aus den Pseudoausprägungen zu den 22 Kontrollnummern, temporär auch NULL.

Diese beiden Relationen enthalten die als transitive Hüllen über die Automatch-Ergebnisse gebildeten Gruppen „ähnlicher“ Meldungen. Innerhalb jeder Gruppe ähnlicher Meldungen werden gleiche Kontrollnummern durch gleiche Pseudoausprägungen gekennzeichnet, um die visuelle Nachbearbeitung zu unterstützen.

Zur besonderen Berücksichtigung von „NULL“-Werten in Kontrollnummern, also von Kontrollnummern zu Patientenattributen mit unbekanntem Wert, wird eine interne Hilfsrelation geführt, die zum bei der Kontrollnummerengenerierung verwendeten Schlüssel den Wert der „leeren“ Kontrollnummer enthält.

Relation Leere_Kontrollnummern

Attribut	Typ	Ausprägungen / Kommentar
<u>key_id</u>	char(20)	Schlüssel zur Kontrollnummerengenerierung.
wert	char(40)	Wert der „leeren“ Kontrollnummer.

Patienten, die im Zuge des Datenabgleichs aufgrund neuer Meldungen neu in die Datenbank aufgenommen werden oder deren Daten sich in Folge von Tumorabgleich bzw. Best-of-Generierung ändern, werden temporär für eine Neuberechnung der Relation *Erweiterte_Tumordaten* notiert.

Relation Geaenderte_Patienten

Attribut	Typ	Ausprägungen / Kommentar
<u>patient_id</u>	number(7)	Schlüssel aus Patienten.

Die Relationen **Meldung-** und **Tumormeldung_Unklarheiten** (zur Spezifikation von Rückfragen), **Anfrage_Kandidaten**, **Loeschanweisungen** (für die Bearbeitung von Auskünften an Patienten oder Einwilligungswiderrufen) sowie **Arzt-** und **Sammelmelder_Totenschein_Infos** (zur Rückmeldung des Vitalstatus an Melder) sind exakt wie in der Vertrauensstelle definiert (s. Abschnitt 1.2.7). Im Falle der Unklarheitsrelationen ist zu beachten, daß – da die entsprechenden

Meldungen in der Registerstelle gelöscht werden – hier zeitweise die referentielle Integrität nicht gewährleistet ist (also keine Verwendung von Fremdschlüsseln!). Ergebnisse der Nachfragen werden ggf. aus den Einträgen, die aus der Vertrauensstelle übermittelt werden, in den entsprechenden Kommentar-Relationen (vgl. Abschnitt 1.2.3) abgelegt. Anschließend werden die Unklarheitstupel gelöscht.

Zusätzlich wird über die vorgenommenen Löschungen von Meldungen (aufgrund durch Patienten widerrufenen Einwilligungen) Buch geführt. (Auskünfte werden in der Registerstelle nicht protokolliert.) Der Verweis auf die Person dient nur dazu, zu erkennen, wie viele unterschiedliche Personen von den Löschungen betroffen sind:

Relation Loeschungen

Attribut	Typ	Ausprägungen / Kommentar
<u>meldung_id</u>	number(8)	Schlüssel gemäß Meldungen – die gelöschte (jetzt also nicht mehr vorhandene) Meldung. Hier ist die referentielle Integrität somit nicht gewahrt (kein Fremdschlüssel!).
patient_ref	number(7)	„Fremdschlüssel“ aus Patienten. Auch hier keine referentielle Integrität. Evtl. auch NULL.
loeschdatum	date	Datum des Löschens.

2.3 Indizes

Zur Steigerung der Effizienz häufiger und komplexerer Anfragen wurden einige Indizes auf dem Schema definiert.²⁸ Auf den epidemiologischen Daten sind dies lediglich Indizes auf der Relation *Erweiterte_Tumordaten* zur schnellen Selektion von Fallzahlen – und zwar auf den Attributkombinationen:

- Wohnort, Diagnosejahr
- Sterbewohnort, Sterbejahr
- Diagnosejahr, Diagnose
- Sterbejahr, Diagnose

Zur Unterstützung des Abgleichs gibt es folgende Indizes über jeweils ein einzelnes Attribut (angeführt sind jeweils Tabellename und alle in der Tabelle indizierten Attribute):

- *Abgleich_Info_PDB*: Kandidat-ID, Geschlecht, Geburtsjahr, Geburtsmonat, Gemeindekennziffer, Treffer
- *Abgleich_Info_KN*: Meldung-ID, Kn-ID, Wert
- *Automatch_Ergebnisse*: Meldung-ID, Meldung-Ref, Position, Bearbeitungsstatus
- *Automatch_Huellen*: Bearbeitungsstatus
- *Automatch_Huellen_Elemente*: Huelle-Ref
- *Kontrollnummeraggregate*: Meldung-ID, Zeile

²⁸ Per Default sind für alle Relationen jeweils über die Primärschlüssel Indizes definiert.

- Meldungen: Meldungsstatus, Meldemodus
- Tumormeldungen: Meldung-ID, Tumor-Ref
- Pathologentumormeldungen: Meldung-ID
- Tumoren: Patient-Ref

3. Unterschiede zum Standard der Arbeitsgemeinschaft bevölkerungsbezogener Krebsregister

Das Datenschema für das Epidemiologische Krebsregister Niedersachsen in der Erprobungsphase basiert im wesentlichen auf dem „Vorschlag für Dokumentationsstandards in epidemiologischen Krebsregistern nach dem Krebsregistergesetz (KRG)“ der GMDS-Projektgruppe „Krebsregister“ vom 24.3.1995 [SAW95] bzw. dessen durch die Arbeitsgemeinschaft bevölkerungsbezogener Krebsregister in Deutschland überarbeiteter Version vom 16.2.1996 [Sch96]. In einzelnen Punkten weicht es im epidemiologischen Datensatz leicht hiervon ab, wobei jedoch stets darauf geachtet wurde, eine Konvertierungsmöglichkeit auf die Dokumentationsstandards (für den bundesweiten Datenabgleich sowie den Datentransfer an das Robert-Koch-Institut zur länderübergreifenden Auswertung in einem auf [Sch96] aufbauenden Format) zu gewährleisten. Im folgenden werden diese Abweichungen von der überarbeiteten Fassung [Sch96] kurz angeführt und erläutert.

3.1 Grundsätzliche Überlegungen

Die Kodierung der Ausprägungen nominaler Attribute mit kleinem Wertebereich (Aufzählungstypen) erfolgt i.a. numerisch als Ziffer von 1 bis 9. Es werden keine mnemotechnischen Codes zur Repräsentation von Werten in der Datenbank verwendet, um eine einfachere und flexiblere Handhabung zu ermöglichen. Zur Unterstützung des Datenaustauschs mit anderen Systemen werden in zu allen derartigen Attributen vorliegenden Kodierungstabellen ggf. Kodierungsschemata abgelegt, die auch Klartextangaben umfassen. Mnemotechnische Codes sind dort jeweils zusätzlich zu numerischen Codes, insb. zur Erleichterung von Dateneingabe und –abgleich, definiert.

3.2 Personenbezogene Angaben

Die Kodierung der *Staatsangehörigkeit* erfolgt nach der Klassifikation des Statistischen Bundesamtes und ist somit deutlich feiner als in den Dokumentationsstandards („deutsch“ bzw. „nicht deutsch“).

Es werden ein- und zweieiige *Mehrlinge* unterschieden.

Angaben zur *Tätigkeit* sollen vierstellig (Dokumentationsstandards: dreistellig) erfaßt werden. Zusätzlich werden Klartexte gespeichert. Die *Dauer* der Tätigkeiten wird jeweils in Monaten anstelle von Jahren angegeben. Angaben aus den verschiedenen Meldungen werden gesammelt, jedoch nicht zu einer aussagekräftigen Tätigkeitsanamnese verdichtet, also nicht miteinander abgeglichen.

Der *Raucherstatus* soll zusätzlich erhoben werden. Hier sind jedoch – insbesondere im Hinblick auf mögliche Ausprägungen dieses Attributs – noch Erfahrungen zu sammeln. Weiterhin werden zusätzlich eine Familien- und eine detailliertere Wohnortanamnese vorgenommen.

Für Pathologenmeldungen kann aus Gründen des Datenschutzes nur eine aggregierte Wohnortangabe²⁹ anstelle der *Gemeindekennziffer* kodiert werden. Entsprechend sind auch keine personenbezogenen *Gauß-Krüger-Koordinaten* verfügbar. Außerdem muß hier auf *Geburtsmonat*, *Staatsangehörigkeit* und *Mehrlingseigenschaft* verzichtet werden.

3.3 Tumorbezogene Angaben

Es besteht die Möglichkeit, Angaben zu mehreren Tumoren als zu einer Erkrankung gehörig zu kennzeichnen. Somit kann insbesondere ein Tumor mehrere Morphologiekodes erhalten. Inwiefern dies benötigt wird, ist noch zu evaluieren. Aus diesem Grunde wurde eine entsprechende Kodierung auch bisher nur auf Meldungsebene vorgesehen – auf Tumorebene steht jeder Tumor (mit *einem* ICD- sowie *einem* ICD-O-Kode) für sich, so daß hier – ebenso wie bei einer Konvertierung auf die Dokumentationsstandards – ggf. die Auswahl eines Kodes erfolgen muß.

Klartexte zur Diagnose (ICD, ICD-O und TNM) werden nicht einzeln (über ein standardisiertes Vokabular) angelegt, sondern gemeinsam in einem Originaltext abgespeichert.

Der *Diagnoseanlaß* (Beschwerden, Früherkennung, Arbeitsmed. Untersuchung, Nachsorge-Untersuchung, Zufallsbefund (bei Autopsie), Sonstiges) wird zusätzlich erfaßt.

Als *Diagnosesicherung* kann zusätzlich „spezielle Diagnostik“ (als Spezialisierung von „klinisch“) kodiert werden.

Die *Dignität* wird getrennt vom ICD-O-Morphologie-Schlüssel abgelegt.

Neben der *TNM* werden keine weiteren Klassifikationsschemata unterstützt (vom Vorschlag der ABKD auch offengehalten). Angaben finden sich ggf. im originalgetreu abgespeicherten Klartext der Meldung (s.o.). Auch für die TNM-Kodierung wird keine einheitliche Syntax vorgegeben. Diesbezüglich sollen noch Erfahrungen gesammelt werden.

Beim Grading wird zusätzlich „medium grade“ (G2-3) angeboten, um entsprechend gemeldete Angaben in die Datenbank aufnehmen zu können. Diese Ausprägung wird ggf. (wie in [EWLP95] empfohlen) auf den höheren Grad G3 konvertiert.

Die Angabe *früherer Tumorleiden* beschränkt sich – gemäß einer Option des ABKD-Vorschlags – auf die Unterscheidung nach erstem, zweitem, weiterem Tumor sowie die Angaben „früheres Tumorleiden vorhanden“ und „Metastase, Primärtumor unbekannt“. Zur Erfassung von Art und Zeitpunkt weiterer Tumorleiden können ggf. zusätzliche Meldungen „generiert“ werden.

²⁹ Dies ist derzeit die Kreiskennziffer – eine differenziertere Lösung in Form einer Definition regionaler Beobachtungseinheiten wird diskutiert. Eine Aggregation auf Kreisebene, d.h. die für Datentransfers an das RKI bzw. den bundesweiten Abgleich vorgesehene Granularität, ist stets möglich.

3.4 Angaben zur Therapie

Zu jeder möglichen *Therapieart* kann als Differenzierung von „unbekannt“ angegeben werden, ob diese definitiv nicht durchgeführt wurde, noch vorgesehen ist oder abgelehnt wurde.

Die Unterscheidung eines kurativen und palliativen *Therapieziels* ist nicht nur für die Operation möglich.

Neben Operation, Bestrahlung und Chemotherapie werden als *Therapieart* auch Hormon- und Immuntherapie sowie Knochenmarkstransplantation angeboten.

3.5 Angaben zum Sterbefall

Die *unmittelbare Todesursache* sowie die verschiedenen gemeldeten *Grundleiden* werden nur unter der Meldung abgelegt. Auswertungen erfolgen zunächst über alle Tumoren, die den Tod bedingt haben (i.a. ist dies das Grundleiden vom Totenschein).

Falls der *Tod tumorbedingt* ist, wird auch erfaßt, welcher Tumor zum Tod ursächlich beigetragen hat (evtl. auch mehrere).

Weiterhin können zusätzlich beliebige Tumore aus der Ursachenkette auf dem Totenschein, ggf. mit dem Zeitraum zwischen Erkrankung und Tod, erfaßt werden.

4. Geodaten in der Registerstelle

Nachfolgend wird die Modellierung der Geodaten in der Registerstelle des EKN näher vorgestellt. Diese lehnt sich an die Modellierung der Daten im Amtlich Topographisch-Kartographischen Informationssystem (ATKIS) der Landesvermessungsämter [AdV91] als unsere Hauptquelle für Geodaten an.

Geodaten werden weitgehend unabhängig von den epidemiologischen Daten verwaltet. Lediglich in der Relation *Erweiterte_Tumordaten* (s. Abschnitt 2.2.7) wird aus dem epidemiologischen Schema auf eine „interne“ Gebietsnummer aus dem Geo-Schema referenziert. Die Verknüpfung erfolgt außerdem über die Kodierung des Wohnortes mit Hilfe von Gemeindekennziffern (bzw. später regionalen Beobachtungseinheiten). Darüber hinaus wurden die Kodierungen von Jahr, Alter und Geschlecht (im Fall der Bevölkerungsdaten des Geo-Schemas) aufeinander abgestimmt. In Abschnitt 4.2.5 wird das Schema der Standardbevölkerungen und der Fallzahlen zu diesen Standardbevölkerungen beschrieben.

4.1 Konzeptionelles ER-Schema

Das Geo-Schema gliedert sich in die Modellierung allgemeiner Geometrie-Objekte, wie Punkte, Linien und Polygone, sowie die Beschreibung konkreter raumbezogener Daten, wie Verwaltungsgebiete, Straßen und Gewässer, in Untersuchungsgebieten. Das konzeptionelle ER-Schema für die Geometrie-Objekte ist in Abbildung 4.1 dargestellt.³⁰ Es ermöglicht den Verzicht auf jegliche redundante Darstellung von Geometrieinformationen. Jedes Geometrieobjekt muß nur einmal abgespeichert sein und kann dann Teil mehrerer Geo-Objekte sein.

Ein Geo-Objekt besteht aus einem Mittelpunkt und Objektteilen. Jedes dieser Objektteile ist entweder ein Punkt-OT (-Objektteil), ein Linien-OT oder ein Flächen-OT. Punkt-OT sind Punkte, Linien-OT Segmente und Flächen-OT Polygone.

Unter Polygonen werden geschlossene Polygonzüge verstanden. Sie werden zur Darstellung von einfachen Flächen genutzt. Daher können sie sowohl Objektteil eines oder mehrerer (flächenförmiger) Geo-Objekte sein, als auch ein Loch in einem oder mehreren solcher Geo-Objekte darstellen. Polygone bestehen aus einem oder mehreren Segmenten.

Segmente sind offene oder geschlossene Polygonzüge, die aus einem Anfangs-, einem End- und evtl. mehreren Stützpunkten bestehen. Auf die gesonderte Darstellung von Anfangs- und Endpunkt wird verzichtet. Ein Segment besteht somit aus Punkten. Die Extension dieses ER-Diagramms muß für die Punkte (über die Punktzuordnung) eine Reihenfolge festlegen. Ein Segment kann Teil eines oder mehrerer Polygone sein. Auch hier sollte in der Extension eine Rei-

³⁰ In der Implementierung dieses Schemas wurden zur Vereinfachung des Datenzugriffs einige Umstrukturierungen vorgenommen. So wurden insbesondere die schraffierten Relationen in die übrigen Entitäten integriert und fallen somit weg.

henfolge festgelegt sein (über Segmentzuordnung). Ein Segment kann außerdem Linien-OT eines oder mehrerer Geo-Objekte sein.

Punkte können Mittelpunkte und/oder Objektteil eines oder mehrerer Geo-Objekte sein. Ebenso können sie Teil eines oder mehrerer Segmente sein. Zusätzlich (siehe Abb. 4.3) können sie noch Mittelpunkte eines oder mehrerer Staßenabschnitte sein.

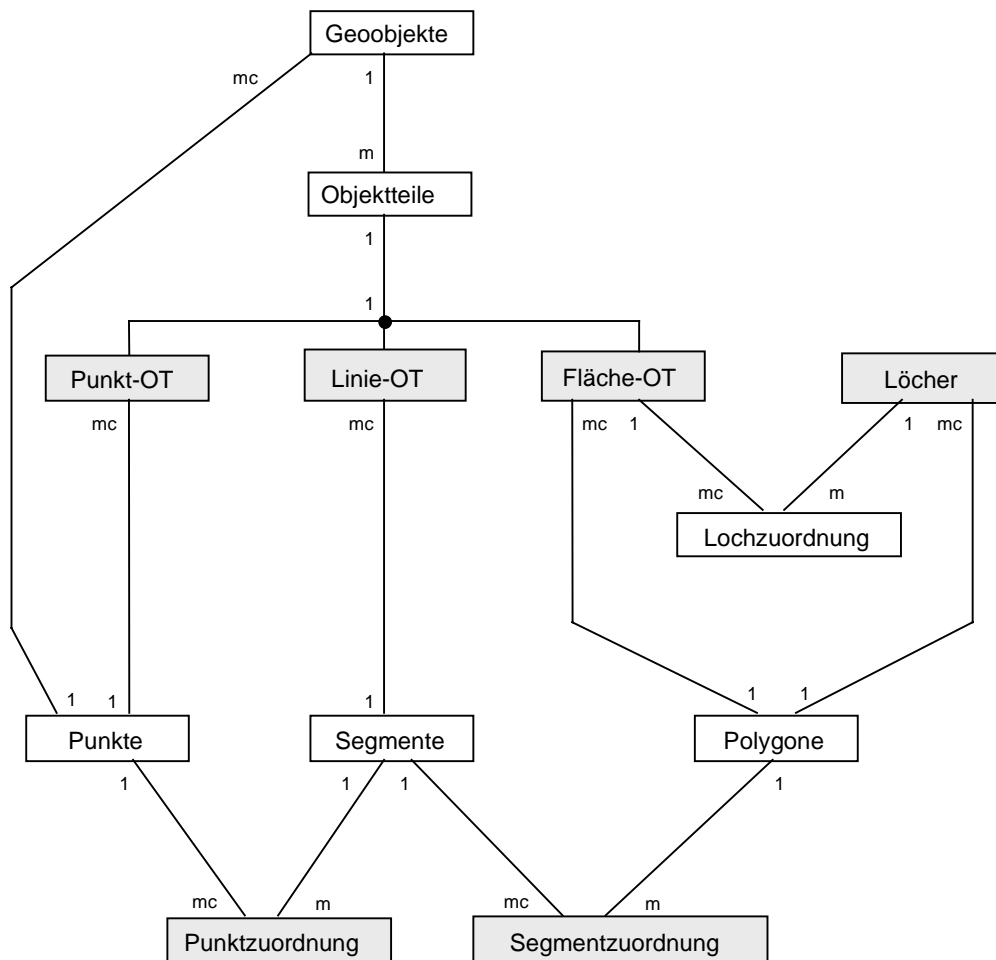


Abb. 4.1: Konzeptionelles ER-Schema für Geometrieobjekte

Als besondere Darstellungsformen sind zu nennen:

- **Flächen:** Flächen (als Geo-Objekte) bestehen aus Einzelflächen (Objektteile, Flächen-OT), von denen jedes Löcher haben kann. Sowohl Einzelflächen als auch Löcher sind Polygone, ein Polygon kann Einzelfläche eines Geo-Objekts und Loch eines anderen Geo-Objekts sein.
- **Straßen:** Im allgemeinen wird eine Straße als offener Polygonzug dargestellt. Für Straßengabelungen u.ä. muß jedoch modelliert werden können, daß mehrere solcher Polygonzüge eine Straße bilden. Also besteht eine Straße (als Geo-Objekt) aus ein oder mehreren Segmenten.

Abbildung 4.2 zeigt das ER-Schema für raumbezogene Daten in Untersuchungsgebieten.³¹ Die beiden wichtigsten Entitäten sind hier Untersuchungsgebiete und Gebiete.

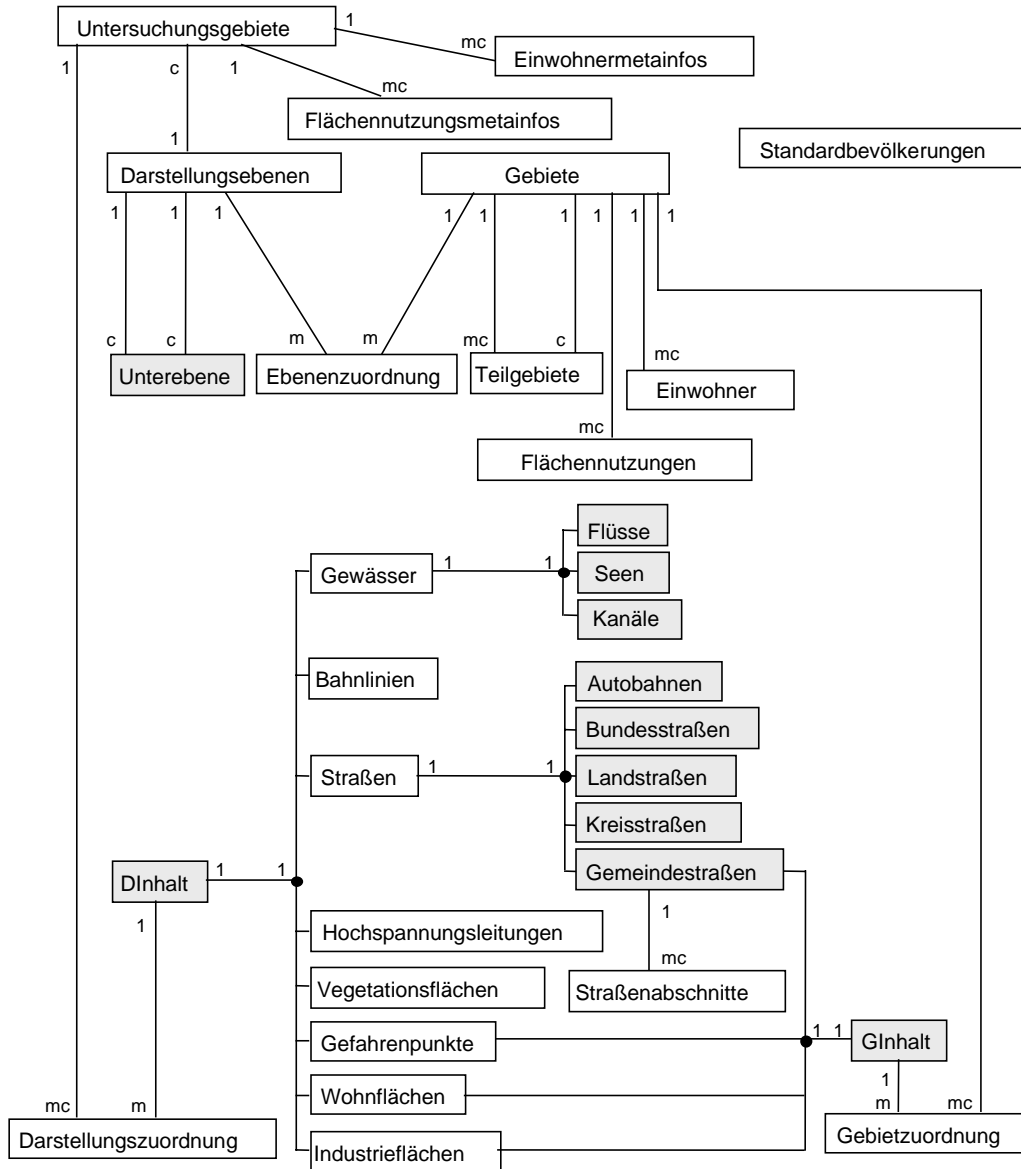


Abb. 4.2: Konzeptionelles ER-Schema für Untersuchungsgebiete

Ein Gebiet ist eine Einheit, für die Auswertungen durchgeführt werden können, z.B. Verwaltungsgebiete wie Regierungsbezirke, Kreise, kreisfreie Städte und Gemeinden. Ein Gebiet kann aus Teilgebieten bestehen. So setzen sich etwa Regierungsbezirke aus Kreisen und kreisfreien Städten oder Städte aus Stadtteilen zusammen. Entsprechend kann ein Gebiet auch Teilgebiet eines anderen Gebiets sein. Weiterhin können einem Gebiet Informationen über Flächennutzungen sowie Bevölkerungsdaten zugeordnet sein. Diese Art der Regionalinformationen ist für alle

³¹ Wieder sind in der Implementierung weggefallene Relationen schraffiert.

Gebiete eines Untersuchungsgebietes gleich und wird in den Flächennutzungs- und Einwohnermetainfos beschrieben.

Gebiete werden über die Entität Ebenenzuordnung zu Darstellungsebenen zusammengefaßt. Eine Darstellungsebene besteht dabei aus den Gebieten, die zusammen auf einer Karte angezeigt werden sollen, also z.B. aus Kreisen und kreisfreien Städten für die Darstellungsebene "Niedersachsen auf Kreisebene". Jede Darstellungsebene kann eine Unterebene haben bzw. Unterebene einer anderen Darstellungsebene sein. Ein Untersuchungsgebiet ist eine ausgezeichnete Darstellungsebene. Als Untersuchungsgebiet können z.B. Niedersachsen oder Weser-Ems definiert sein, wobei die zugehörige Darstellungsebene jeweils die "höchste" ist (Niedersachsen auf Landesebene, Weser-Ems auf Regierungsbzirksebene usw.).

Parallel zu den Gebieten der Untersuchungsgebiete werden noch Bevölkerungsdaten zu ausgewählten Standardbevölkerungen (Welt, Europa, BRD) verwaltet.

Die Entitäten Gewässer, Hochspannungsleitungen, Straßen, Bahnlinien, Vegetationsflächen, Industrieflächen, Wohnflächen und Gefahrenpunkte dienen zwei Zwecken. Zum einen sollen sie auf einer Karte dargestellt werden können, um die Orientierung im jeweiligen Untersuchungsgebiet zu erleichtern. Daher sind sie über DInhalt und Darstellungszuordnung den Untersuchungsgebieten zugeordnet. Einige dieser Objekte werden jedoch auch für interne Berechnungen benötigt, wie z.B. Gemeindestraßen zur Adreßumsetzung, Wohnflächen für die Aufteilung der Bevölkerung einer Gemeinde auf kleinere Gebiete usw. Daher werden einige dieser Entitäten über GInhalt und Gebietzuordnung der Entität Gebiet zugeordnet.

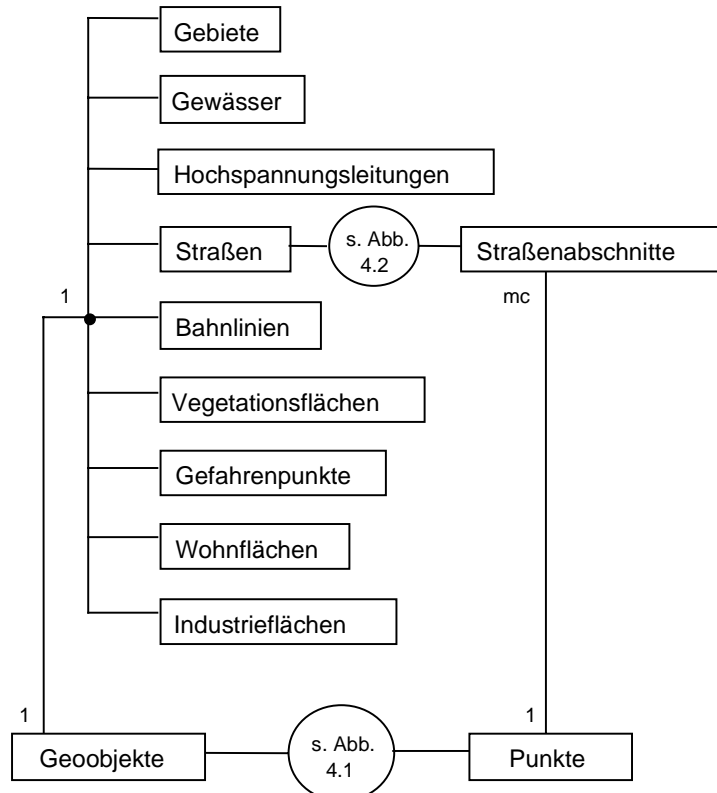


Abb. 4.3: Beziehungen zwischen Gebieten und Geo-Objekten

Gemeindestraßen können in Straßenabschnitte aufgeteilt sein. Dieses wird notwendig, wenn für lange Straßen eine genauere Adreßumsetzung geschehen soll. Straßenabschnitte können nicht gesondert dargestellt werden (sie sind kein Geo-Objekt), haben jedoch einen eigenen Mittelpunkt.

Abbildung 4.3 zeigt den Zusammenhang zwischen den Entitäten aus Abbildung 4.2 und den Geo-Objekten (vgl. Abbildung 4.1).

4.2 Relationen

Das im vorangegangenen Abschnitt vorgestellte Datenschema entspricht der dritten Normalform. Für eine Umsetzung in eine reale Datenbank wäre die Modellierung sämtlicher vorgestellter Entitäten jedoch in Bezug auf die Laufzeit ineffizient. Daher wird in diesem Abschnitt eine Modellierung in Tabellenform einer relationalen Datenbank (ORACLE) vorgestellt, die einen effizienteren Zugriff zuläßt. Wie in Abschnitt 4.1 bzw. Abbildung 4.1 und 4.2 angedeutet, entfallen hierdurch verschiedene Relationen. Das so entstehende Datenschema entspricht nicht mehr der dritten Normalform.

4.2.1 Verwendete Domänen

In der folgenden Tabelle sind die für das Datenschema benötigten Domänen aufgeführt.

Domänenname	Basistyp
Altersangabe	number(3)
Codenummer	number(4)
StaDInhaltart	{Gewässer, Hochspannungsleitung, Straße, Bahnlinie, Vegetationsfläche, Wohnfläche, Industriefläche, Gefahrenpunkt}
Ebenenummer	number(10)
Einheitentyp	{ha}
Einwohnernummer	number(10)
Erhebungsjahr	number(4)
Flächenanteil	number(4,2)
Flächenart	{Gesamtfläche, Gebäude- und Freifläche, Wohnfläche, Gewerbe- und Industriefläche, Betriebsfläche, Abbauland, Erholungsfläche, Grünanlage, Verkehrsfläche, Straße / Weg / Platz, Landwirtschaftsfläche, Moor, Heide, Waldfläche, Wasserfläche, Flächen anderer Nutzung, Unland}
Flächenzahl	number(13,3)
Gebietnummer	number(10)
Gefahrenpunktart	{Bergbaubetrieb, Abfalldeponie, Raffinerie, Kraftwerk, Umspannwerk, Kläranlage, Heizwerk, Abfallbeseitigungsanlage, Flughafen}

Gefahrenpunktkennzeichnung	{Erde, Ton, Lehm, Löß, Torf, Gestein, Lava, Mergel, Bims, Quarz, Salz, Schiefer, Tuff, Kalkstein, Granit, Steinkohle, Braunkohle, Kies, Sand, Erz, Eisen, Kupfer, Blei, Silber, Gold, Uran, Zink, sonstige} ∪ {Hausmüll, Industriemüll, Sondermüll, sonstiges} ∪ {Wasser, Kernkraft, Sonne, Wind, Gezeiten, Verbrennung, Kohle, Öl, Gas, Müll, sonstige} ∪ {Sonne, Verbrennung, Kohle, Öl, Gas, Müll, Erdwärme, sonstige} ∪ {Hausmüll, Industriemüll, Sondermüll, sonstige} ∪ {international, regional, militärisch}
Gemeindekennziffer	number(10)
Geoname	char(33)
Geometriety	{Punkt, Segment, Polygon }
Geschlechtskennung	<s. Kapitel 1 und 2>
Gewässerart	{Fluß, Kanal, See}
GIInhaltart	{Wohnfläche, Industriefläche, Gemeindestraße, Gefahrenpunkt }
Hausnummer	number(4)
Industrieart	{Industrie_und_Gewerbe, Einkaufszentrum, Lager_und_Abstellflächen, Handel_und_Banken, Dienstleistung, Versorgung, Entsorgung, Kommunikation, Sonstige}(aus Atkis)
Inhaltnummer	number(10)
Koordinate	number(10)
Objektnummer	number(10)
Objektteilnummer	char(3)
Punktkenung	{P, V}
Quellenbezeichnung	{Atkis, Aktive_Ineks, NLVA}
Quellenschlüssel	char(20)
Reihenfolgennummer	number(4)
Richtung	{normal, umgekehrt}
Standardgebiet	{BRD, Europa, Welt, W35-64}
Straßenwidmung	{Autobahn, Bundesstraße, Landesstraße, Kreisstraße, Gemeindestraße}
Straßenteilnummer	number(10)
Teilnummer	number(10)
Vegetationsflächenart	{Wald_und_Grünflächen, Landwirtschaft}
Verwaltungsart	{Land, Bundesland, Regierungsbezirk, Kreis, Kreisfreie Stadt, Samtgemeinde, Gemeinde, Gemeindefreies Gebiet}
Zählart	{gerade, ungerade, vollstaendig}

4.2.2 Tabellen

In diesem Abschnitt werden die verwendeten Datenbanktabellen sowie für jede Tabelle einzuhaltende Konsistenzbedingungen beschrieben.

Folgende Konsistenzbedingungen gelten für alle Tabellen:

- KA 1: Unterstrichene Attribute sind Primärschlüsselattribute, die eindeutig (tupelidentifizierend) sein müssen.
- KA 2: Referentielle Integrität muß eingehalten werden, d.h. wenn in einer Extension dieses Schemas ein Tupel in einer Tabelle A einen Fremdschlüssel aus einer Tabelle B enthält, so muß es in B ein Tupel mit diesem Schlüssel als Primärschlüssel geben.

Soweit nicht anders angegeben, sind keine NULL-Werte erlaubt.

Relation Geoobjekte

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>objnr</u>	number(10)	Objektnummer	Eindeutiger Schlüssel.
mittelpunkt	number(10)	Teilnummer	Fremdschlüssel aus Punkte.
datenquelle	char(15)	Quellenbezeichnung	Die Herkunft der Geometriedaten.
qschluessel	char(20)	Quellenschlüssel	In einen String umgewandelter Schlüssel dieses Geoobjekts in der Datenquelle. Diese Angabe wird benötigt, um bei einem Update (z.B. in ATKIS) die Daten zu reidentifizieren.

KO 1: Jedes Geoobjekt wird in genau einer der Relationen Gebiet, Gewaesser, Hochspannungsleitungen, Straße, Bahnlinien, Vegetationsflaechen, Wohnflaechen, Industrieflaechen oder Gefahrenpunkte referenziert.

KO 2: Wann immer sich zwei Geoobjekte schneiden, gilt: Der Schnittpunkt ist in der Relation Punkte vorhanden und hat als PunktArt Vertex.

Relation Objektteile

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>objnr</u>	number(10)	Objektnummer	Schlüssel aus Geoobjekte.
<u>objteilnr</u>	char(3)	Objektteilnummer	Bildet zusammen mit der Objektnummer den Primärschlüssel für die Objektteile.
objektteil	number(10)	Teilnummer	Je nach ObjGeotyp Fremdschlüssel aus Punkte, Segmente oder Polygone.
objgeotyp	char(8)	Geometrietyp	Diskriminierendes Attribut für die Spezialisierung der Objektteile.
qschluessel	char(20)	Quellenschlüssel	<s. Geoobjekte>

Zu jedem Geoobjekt gehören ein oder mehrere Teilobjekte. Diese sind disjunkt spezialisiert in Punkte, Linien (= Segmente) und Flächen. Diese Spezialisierung wird hier durch ein diskriminierendes Attribut anstatt durch Entitäten modelliert. Die in Abbildung 4.1 gezeigten Entitäten Punkt-OT, Linie-OT und Fläche-OT fallen daher weg.

KO 3: Jede ObjNr aus Geoobjekte (d.h. jedes Geo-Objekt) wird mindestens einmal referenziert.

Relation Segmente

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>segmentnr</u>	number(10)	Teilnummer	Jedes Segment erhält eine eindeutige Nummer.
<u>punktnr</u>	number(4)	Reihenfolgenummer	Die Stützpunkte eines Segments sind von 0 an durchnummeriert.
stuetzpunkt	number(10)	Teilnummer	Fremdschlüssel aus Punkte.

Die Entität Punktzuordnung wird nicht modelliert, sondern die Zugehörigkeit von Punkten zu Segmenten wird direkt in der Relation Segmente dargestellt.

KS 1: Punktnummer muß für jedes Segment von 0 bis zu einer Obergrenze durchnummeriert sein.

KS 2: Für je zwei Segmente gilt, daß sie sich in mindestens einem Punkt und/oder in der Reihenfolge der Punkte unterscheiden. (Reihenfolge ist für diese Bedingung unabhängig von der Richtung, d.h. ein Segment darf nicht einmal von Endpunkt A nach Endpunkt B und zusätzlich von Endpunkt B nach Endpunkt A gespeichert sein).

KS 3: Jedes Segment wird entweder in Objektteile oder in Polygone referenziert.

KS 4: Sei für ein Segment s die höchste Reihenfolgenummer s_n . Dann haben die Stützpunkte in der Relation Punkte, die in der Relation Segmente für ein Segment referenziert werden, folgende PunktArt: Für die Punkte mit der Reihenfolgenummer 0 und s_n (V)ertex und für alle Punkte mit Reihenfolgenummer >0 und Reihenfolgenummer $<s_n$ (P)unkt.

KS 4 ergibt zusammen mit KG 2 und der unten stehenden Bedingung KP 1 folgende Eigenschaft: Segmente schneiden sich nicht, sie berühren sich höchstens an ihren Endpunkten. Damit wird die Bestimmung von Schnittpunkten von Geoobjekten stark vereinfacht, da diese entweder Punkte sind oder aus Segmenten bestehen. Als Schnittpunkte kommen somit im ersten Fall nur der Punkt selber und im zweiten Fall nur Vertices in Frage.

Relation Punkte

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>punktnr</u>	number(10)	Teilnummer	Jeder Koordinatenpunkt erhält eine eindeutige Nummer.
punktrechts	number(10)	Koordinate	Gauß-Krüger-Koordinate.
punkthoch	number(10)	Koordinate	dito
punktart	char(1)	Punktbezeichnung	(P)unkt oder (V)ertex.

Die Gauß-Krüger-Koordinaten der Punkte beziehen sich alle auf den 9. Längengrad. Dazu ist eine Umwandlung der Koordinaten aus den Datenquellen nötig. Auf die dazu benötigten Formeln wird hier nicht näher eingegangen.

KP 1: Die Koordinaten jedes Punktes sind von denen aller anderen Punkte verschieden.

KP 2: Jeder Punkt wird in mindestens einer der Relationen Geoobjekte, Objektteile, Segmente, oder Straßenabschnitte referenziert.

Relation Polygone

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>polygonnr</u>	number(10)	Teilnummer	Jedes Polygon erhält eine eindeutige Nummer.
<u>segmentnr</u>	number(4)	Reihenfolgenummer	Die Segmente, aus denen ein Polygon besteht, sind von 0 an durchnummeriert.
segment	number(19)	Teilnummer	Repräsentiert ein Segment mit allen Punkten aus der Relation Segmente.
orientierung	char(10)	Richtung	Gibt die Orientierung des betreffenden Segments an.

Die Entität Segmentzuordnung aus Abbildung 4.1 kann wegfallen, da die entsprechenden Informationen direkt in der Relation Polygone dargestellt werden kann.

KF 1: Für jedes Polygon ist Segmentnummer von 0 bis zu einer Obergrenze durchgehend vorhanden.

- KF 2: Je zwei Polygone unterscheiden sich mindestens in einem Segment.
- KF 3: Zwei Segmente eines Polygons, deren Segmentnummer (Reihenfolgenummer) aufeinander folgt, besitzen den gleichen Anfangs- und/oder Endpunkt. Alle Segmente eines Polygons bilden einen geschlossenen Linienzug.
- KF 4: Jedes Polygon wird in Objektteile referenziert.

Relation Lochzuordnung

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>objnr</u>	number(10)	Objektnummer	Fremdschlüssel aus Objektteile.
<u>objteilnr</u>	char(3)	Objektteilnummer	dito.
<u>loch</u>	number(10)	Teilnummer	Repräsentiert ein Polygon mit allen Segmenten aus der Relation Polygone.

Weder die Entität Loch noch die Entität Fläche-OT aus Abbildung 4.1 wird modelliert. Daher realisiert Lochzuordnung eine Relation zwischen Objektteilen und Polygonen.

- KL 1: Das durch den Fremdschlüssel referenzierte Objektteil muß eine Fläche sein (ObjGeo-Typ).

4.2.3 Untersuchungsgebiete

Zur Beschreibung der administrativen Struktur von Untersuchungsgebieten (etwa Niedersachsen oder Weser-Ems) dienen die folgenden Relationen:

Relation Untersuchungsgebiete

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>untobjnr</u>	number(10)	Gebietnummer	Eindeutige Nummer.
ebenennr	number(10)	Ebenennummer	Fremdschlüssel aus Darstellungsebenen.
name	char(33)	Geoname	Name des Untersuchungsgebiets (Niedersachsen, Weser-Ems,...).
minrechts	number(10)	Koordinate	Untere linke Ecke des Untersuchungsgebiets („Rand“ der Karte) / X-Koordinate.
minhoch	number(10)	Koordinate	dito / Y-Koordinate.
maxrechts	number(10)	Koordinate	Obere rechte Ecke des Untersuchungsgebiets („Rand“ der Karte) / X-Koordinate.
maxhoch	number(10)	Koordinate	dito / Y-Koordinate.

Relation Gebiete

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>gebietnr</u>	number(10)	Gebietnummer	Eindeutige Nummer.
geometrie	number(10)	Objektnummer	Fremdschlüssel aus Geoobjekte.
name	char(33)	Geoname	Der Name des Gebiets wird aus ATKIS übernommen.
gkz	number(10)	Gemeindekennziffer	Falls das Gebiet eine Gemeindekennziffer besitzt, kann hier die GKZ gespeichert werden, ansonsten bleibt das Feld frei (NULL).

art	char(25)	Verwaltungsart	Eine Bezeichnung der Art des Gebiets (Land, Regierungsbezirk usw.).
-----	----------	----------------	---

Relation Teilgebiete

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>hauptgebiet</u>	number(10)	Gebietnummer	Fremdschlüssel aus Gebiete.
<u>teilgebiet</u>	number(10)	Gebietnummer	Fremdschlüssel aus Gebiete.

Relation Darstellungsebenen

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>ebenenr</u>	number(10)	Ebenennummer	Eindeutige Nummer.
unterebene	number(10)	Ebenennummer	Fremdschlüssel aus Darstellungsebenen – ggf. NULL, falls keine Unterebene existiert.
name	char(33)	Geoname	Der Name der Ebene (Kreisebene, Landesebene,...).

Durch das Attribut Unterebene wird eine eigene Entität für diese Beziehung eingespart.

Relation Ebenenzuordnung

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>ebenenr</u>	number(10)	Ebenennummer	Fremdschlüssel aus Darstellungsebenen.
<u>gebiet</u>	number(10)	Gebietnummer	Fremdschlüssel aus Gebiete.

Für die fünf Relationen Untersuchungsgebiete, Gebiete, Teilgebiete, Darstellungsebenen und Ebenenzuordnung gelten folgende Konsistenzbedingungen:

- KG 1: Jede Darstellungsebene wird höchstens einmal in Untersuchungsgebiete referenziert.
- KG 2: Die transitive Hülle der Relation Teilgebiete ist asymmetrisch.
- KG 3: Die transitive Hülle der Relation Darstellungsebenen ist asymmetrisch.
- KG 4: Wenn ein Gebiet A Teilgebiet eines Gebiets B ist, so wird über die Relation Ebenenzuordnung A einer Darstellungsebene A' und B einer Darstellungsebene B' zugeordnet. A' ist dann in der Relation Darstellungsebenen eine Unterebene von B'.
- KG 5: Eine Darstellungsebene A, die in der Relation Untersuchungsgebiete referenziert wird, besitzt keine übergeordnete Ebene, d.h. es gibt keine Darstellungsebene B, so daß A Unterebene von B ist.
- KG 6: Die Koordinaten der Punkte aller GeoObjekte, die einem Untersuchungsgebiet zugeordnet sind, liegen alle innerhalb des durch (MinRechts,MinHoch) und (MaxRechts, MaxHoch) definierten Rechtecks.

Weiterhin können einem Untersuchungsgebiet verschiedene geographische Informationen zugeordnet sein:

Relation Gewaesser

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>wasobjnr</u>	number(10)	Inhaltnummer	Eindeutige Nummer.
geometrie	number(10)	Objektnummer	Fremdschlüssel aus Geoobjekte.
wasart	char(5)	Gewässerart	Die Gewässerart unterteilt die Klasse der Gewässer in mehrere Unterklassen.
wasname	char(33)	Geoname	Der Gewässername ist der in ATKIS erfaßte Name eines Flusses, eines Sees oder eines Kanals.
wasflaeche	number(13,3)	Flächenzahl	Die Fläche wird nur bei flächenförmigen Objekten (Seen) eingetragen.

Die disjunkte Spezialisierung wird wieder durch ein diskriminierendes Attribut anstatt durch Entitäten übernommen.

Relation Hochspannungsleitungen

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>hspobjnr</u>	number(10)	Inhaltnummer	Eindeutige Nummer.
geometrie	number(10)	Objektnummer	Fremdschlüssel aus Geoobjekte.

Relation Strassen

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>strobjnr</u>	number(10)	Inhaltnummer	Eindeutiger Schlüssel.
geometrie	number(10)	Objektnummer	Fremdschlüssel aus Geoobjekte.
strwidmung	char(20)	Straßenwidmung	Diskriminierendes Attribut (Art der Straße).
strbez	char(33)	Geoname	Für Gemeindestraßen der Name der Straße, für andere die Kurzbezeichnung (A1, K232 usw.).

Es wird wieder keine Entität für die disjunkte Spezialisierung modelliert.

KR 1: Eine Gemeindestraße liegt vollständig in einer Gemeinde.

Relation Strassenabschnitte

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>strobjnr</u>	number(10)	Inhaltnummer	Fremdschlüssel aus Straßen
<u>strteilnr</u>	number(10)	Straßenteilnummer	Zusammen mit strobjnr eindeutiger Schlüssel.
mittelpunkt	number(10)	Teilnummer	Fremdschlüssel aus Punkte.
hnrrechtsvon	number(4)	Hausnummer	Die drei Felder geben den Hausnummernbereich des Straßenabschnitts auf der rechten Straßenseite an.
hnrrechtsbis	number(4)	Hausnummer	
hnrrechtszaehl	char(12)	Zählart	
hnrlinksvon	number(4)	Hausnummer	Die drei Felder geben den Hausnummernbereich des Straßenabschnitts auf der linken Straßenseite an.
hnrlinksbis	number(4)	Hausnummer	
hnrlinkszaehl	char(12)	Zählart	

Rechts und Links werden entsprechend der Geometrie der zu der Straße dieses Straßenabschnitts gehörenden Geometrie festgelegt, d.h. für das entsprechende Segment ist die Blickrichtung vom Punkt mit der Punktnummer 0 hin zum Punkt mit der höchsten Nummer.

KR 2: Die Straße, zu der ein Straßenabschnitt gehört, muß eine Gemeindestraße sein.

KR 3: Für die Abschnitte einer Straße sind die durch die Hausnummernbereiche angegebenen Hausnummern disjunkt.

Relation Bahnlinien

Attribut	Typ	Domäne	Ausprägungen / Kommentare
bahnobjnr	number(10)	Inhaltnummer	Eindeutige Nummer.
geometrie	number(10)	Objektnummer	Fremdschlüssel aus Geoobjekte.
bahnbez	char(33)	Geoname	Die Bahnlinienbezeichnung ist eine betriebstechnische Kurzbezeichnung der Bundesbahn.

Wohnflächen, Industrieflächen und Gefahrenpunkte können jeweils Gemeinden zugeordnet sein, Vegetationsflächen können "gemeindeübergreifend" sein. Wird bei der Übernahme von Daten in eine Extension dieses Schemas festgestellt, daß Wohn- oder Industrieflächen über Gemeindegrenzen hinausgehen, so müssen diese Flächen aufgeteilt werden (siehe Konsistenzbedingungen).

Relation Vegetationsflaechen

Attribut	Typ	Domäne	Ausprägungen / Kommentare
vegobjnr	number(10)	Inhaltnummer	Eindeutige Nummer
geometrie	number(10)	Objektnummer	Fremdschlüssel aus Geoobjekte.
vegart	char(25)	Vegetationsflächenart	Die Vegetationsflächenart unterteilt die Klasse der Vegetationsflächen in zwei Unterklassen (s.u.).
vegflaeche	number(13,3)	Flächenzahl	Die Fläche wird berechnet und kann zur Darstellungsklassifizierung benutzt werden.

VegArt unterteilt die Vegetationsflächen in Wald_und_Grünflächen und Landwirtschaft. Zu Wald_und_Grünflächen gehören die ATKIS-Objektarten 2227 (Grünanlage), 4107 (Wald, Forst) und 4108 (Gehölz), zu Landwirtschaft die Objektarten 4101 (Ackerland), 4102 (Grünland), 4103 (Gartenland) und 4109 (Sonderkultur).

Relation Wohnflaechen

Attribut	Typ	Domäne	Ausprägungen / Kommentare
wohobjnr	number(10)	Inhaltnummer	Eindeutige Nummer.
geometrie	number(10)	Objektnummer	Fremdschlüssel aus Geoobjekte.
siedflaeche	number(13,3)	Flächenzahl	Die Fläche wird berechnet und kann zur Darstellungsklassifizierung benutzt werden.

Zu den Wohnflächen gehören die ATKIS-Objektarten 2111 (Wohnbaufläche), 2113 (Fläche gemischter Nutzung) und 2114 (Fläche besonderer funktionaler Prägung).

Relation Industrieflaechen

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>indobjnr</u>	number(10)	Inhaltnummer	Eindeutiger Schlüssel.
geometrie	number(10)	Objektnummer	Fremdschlüssel aus Geoobjekte.
indart	char(30)	Industrieart	Die Industrieart unterteilt die Klasse der Industrieflächen in mehrere Unterklassen.
indflaeche	number(13,3)	Flächenzahl	Die Fläche wird berechnet und kann zur Darstellungsklassifizierung benutzt werden.

Zu den Industrieflächen gehören die ATKIS-Objektarten 2112 (Industrie- und Gewerbeflächen).

Relation Gefahrenpunkte

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>gefobjnr</u>	number(10)	Inhaltnummer	Eindeutiger Schlüssel.
geometrie	number(10)	Objektnummer	Fremdschlüssel aus Geoobjekte.
gefart	char(25)	Gefahrenpunktart	Die Gefahrenpunktart unterteilt die Klasse der Gefahrenpunkte in mehrere Unterklassen.
gefkennung	char(25)	Gef.pkt.kennzeichnung	Die Gefahrenpunktkenzeichnung wird in ATKIS nicht mit erfaßt und muß nachgetragen werden.

Zu den Gefahrenpunkten gehören die durch GefArt bezeichneten ATKIS-Objekte.

Für Wohnflaechen, Industrieflaechen und Gefahrenpunkte gilt:

KI 1: Die durch die Geometrie referenzierten Geo-Objekte liegen vollständig innerhalb der Geometrie jeweils einer Gemeinde.

Relation Gebietzuordnung

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>gebiet</u>	number(10)	Gebietnummer	Fremdschlüssel aus Gebiete.
<u>typ</u>	char(25)	GIInhaltart	Disjunkte Spezialisierung für den Inhalt.
<u>inhalt</u>	number(10)	Inhaltnummer	Je nach Typ Fremdschlüssel von Wohnflächen, Industrieflächen, Straßen oder Gefahrenpunkten.

Die Entity GIInhalt wird nur durch das diskriminierende Attribut in Gebietzuordnung modelliert.

Relation Darstellungszuordnung

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>gebiet</u>	number(10)	Gebietnummer	Fremdschlüssel aus Untersuchungsgebiete.
<u>typ</u>	char(25)	DIInhaltart	Disjunkte Spezialisierung für den Inhalt.

<u>inhalt</u>	number(10)	Inhaltnummer	Je nach Typ Fremdschlüssel von Hochspannungsleitungen, Gewässer, Bahnlinien, Straßen, Vegetationsflächen, Wohnflächen, oder Industrieflächen.
---------------	------------	--------------	---

Die Entity DInhalt wird nur durch das diskriminierende Attribut in Darstellungsordnung modelliert.

4.2.4 Regionalinformationen

Derzeit werden lediglich Einwohnerzahlen und Flächennutzung als Regionalinformationen modelliert.

Einwohnerzahlen werden in CARLOS sowohl für das Geomodul als auch für die Statistik auf den epidemiologischen Daten benötigt. Um diese Zahlen nicht doppelt zu halten, werden sie in einer Tabelle gespeichert, die dann von beiden Modulen genutzt wird. Dies bedeutet, daß die Einwohner für jedes Gebiet in den für die Statistik benötigten Gruppeneinteilungen (5-Jahres-Gruppen) gespeichert werden müssen.

Da die Gruppeneinteilung keine feste ist, werden als Metainformation in einer Tabelle diese Gruppen festgehalten. Diese Tabelle existiert nur einmal pro Untersuchungsgebiet und hat den folgenden Aufbau:

Relation Einwohnermetainfos

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>untobjnr</u>	number(10)	Gebietsnummer	Fremdschlüssel aus Untersuchungsgebiete.
<u>jahr</u>	number(4)	Erhebungsjahr	Jahr der Erhebung der Daten in dieser Altersgruppe.
<u>geschlecht</u>	number(1)	Geschlechtskennung	Geschlecht der Gruppe.
<u>altermin</u>	number(3)	Altersangabe	Mindestalter dieser Gruppe.
altermax	number(3)	Altersangabe	Maximales Alter dieser Gruppe (NULL-Wert für unendlich).

KE 1: Für jedes Untersuchungsgebiet, jedes Geschlecht und jedes Jahr bilden die durch AlterMin und AlterMax definierten Intervalle eine vollständige, disjunkte Überdeckung des Intervalls $[0..∞)$.

Für die Statistik benötigt CARESS die Einwohnerzahlen nur auf der untersten, genauesten Hierarchieebene, die Zahlen für Gebiete höherer Hierarchieebenen werden durch Addition der Teilgebiete auf der untersten Ebene ermittelt. Für die Regionalinformation, die durch das Geomodul ermittelt werden, müssen die Zahlen jedoch auf allen Hierarchieebenen zur Verfügung stehen. Sobald also Regionalinformationen genutzt werden, müssen die Einwohnerzahlen für alle Gebiete vollständig definiert sein.

Relation Einwohner

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>gebiet</u>	number(10)	Gebietsnummer	Fremdschlüssel aus Gebiete.

<u>untobjnr</u>	number(10)	Gebietnummer	Fremdschlüssel aus Einwohnermetainfos.
<u>jahr</u>	number(4)	Erhebungsjahr	dito.
<u>geschlecht</u>	number(1)	Geschlechtskennung	dito.
<u>altermin</u>	number(3)	Altersangabe	dito.
einwohnerzahl	number(10)	Einwohnernummer	Einwohner des Gebiets.

Die UntObjNr muß in diese Relation aufgenommen werden, da ein Gebiet zu mehreren Untersuchungsgebieten gehören kann, innerhalb derer die Gruppeneinteilung verschieden ist.

Für die Flächennutzungen als weitere Regionaldaten existieren ebenfalls Metadaten, die die vorhandenen Arten der Flächennutzung beschreiben. Diese sind wieder den Untersuchungsgebieten zugeordnet. Die Art der Fläche wird kodiert, um in der Tabelle Flächennutzung Speicherplatz zu sparen.

Wenn zu einem Untersuchungsgebiet Flächennutzungsdaten existieren, dann müssen diese unbedingt für alle Gebiete des Untersuchungsgebiets definiert sein, um diese in CARESS zu nutzen. Es reicht nicht aus Flächennutzungsdaten nur auf der feinsten Ebene zu definieren, da CARESS diese für höhere Ebenen nicht summiert. Wenn Flächennutzungsdaten verfügbar sind, so müssen auch die Einwohnerdaten auf den höheren Ebenen des Untersuchungsgebiets verfügbar sein, da bei der Abfrage für die Regionalinformationen die Einwohnerzahlen ebenfalls nicht summiert werden.

Relation Flaechennutzungsmetainfos

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>untobjnr</u>	number(10)	Gebietnummer	Fremdschlüssel aus Untersuchungsgebiete.
<u>code</u>	number(4)	Codenummer	Kodierung dieser Gruppe.
<u>jahr</u>	number(4)	Erhebungsjahr	Jahr der Erhebung der Daten.
<u>art</u>	char(30)	Flächenart	Name der Gruppe zum Kode.
<u>einheit</u>	char(10)	Einheitentyp	Einheit der Flächen (derzeit Hektar).

Die eigentliche Flächennutzungsinformation ist wieder den Gebieten zugeordnet. Wieder muß UntObjNr aufgenommen werden:

Relation Flaechennutzungen

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>gebiet</u>	number(10)	Gebietnummer	Fremdschlüssel aus Gebiete.
<u>untobjnr</u>	number(10)	Gebietnummer	Fremdschlüssel aus Flächennutzungsmetainfos.
<u>code</u>	number(4)	Codenummer	dito.
<u>jahr</u>	number(4)	Erhebungsjahr	dito.
<u>flaeche</u>	number(13,3)	Flächenzahl	Größe der Fläche (berechnet oder aus anderen Datenbeständen übernommen).

4.2.5 Standardbevölkerungen

Externe Standardbevölkerungen ermöglichen es, die Fallzahlen im Untersuchungsgebiet auf die Bevölkerungsstruktur der externen Standardbevölkerung hochzurechnen. So können die Fallzahlen von Gebieten mit unterschiedlicher Bevölkerungsstruktur verglichen werden.

Die Tabelle Standardbevölkerungen weicht vom Muster der Einwohnertabelle ab, da sie weder Untersuchungs- noch anderen Gebieten zugeordnet sind, sondern lediglich standardisierte Bezugsbevölkerungen aufnehmen. Es gibt ebenfalls keine Metainformationen, da die Gebiete, für die Standardbevölkerungen existieren, weder aus Gebieten bestehen noch zu anderen Gebieten gehören.

Relation Standardbevoelkerungen

Attribut	Typ	Domäne	Ausprägungen / Kommentare
<u>bezugsraum</u>	char(6)	Standardgebiet	Der Name der Standardbevölkerung.
<u>geschlecht</u>	number(1)	Geschlechtskennung	Das Geschlecht dieser Gruppe.
<u>altermin</u>	number(3)	Altersangabe	Das minimale Alter dieser Gruppe.
<u>altermax</u>	number(3)	Altersangabe	Das maximale Alter dieser Gruppe.
<u>einwohnerzahl</u>	number(10)	Einwohnernummer	Größe der Gruppe.

SB 1: Für jeden Bezugsraum und jedes Geschlecht bilden die durch AlterMin und AlterMax definierten Intervalle eine vollständige, disjunkte Überdeckung des Intervalls $[0.. \infty)$.

Zu einigen Standardbevölkerungen werden zusätzlich (externe) Fallzahlen abgelegt, die zur Berechnung von erwarteten Raten im Untersuchungsgebiet benutzt werden können. Diese Raten rechnen die Risiken in einer externen Standardbevölkerung auf die Bevölkerung im Untersuchungsgebiet hoch.

Immer wenn Fallzahlen aus einer externen Quelle gespeichert werden sollen (in der Relation Externe_Fallzahlen), muß diesen Fallzahlen eine Standardbevölkerung (Relation Standardbevoelkerung) zugeordnet sein. Für die Fallzahlen wird in der Relation Externe_Fallzahlen_Metainfos gespeichert, nach welchen Attributen die zugehörigen Fallzahlen unterteilt sind. Für jedes dieser Attribute ist dessen Wertebereich in den Relationen Extern_Diagnose, usw. gespeichert (siehe Abbildung.4.4).

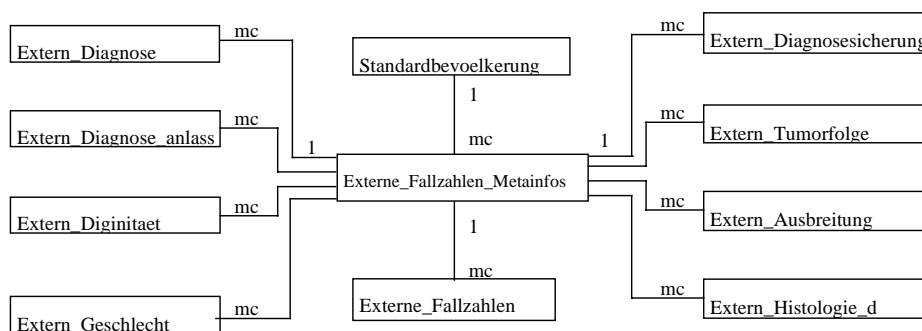


Abb. 4.4: Standardbevolkerungen mit Fallzahlen

Zu jedem externe Standard existiert genau ein Tupel in der Relation Externe_Fallzahlen_Metainfo. Dieses besitzt eine eindeutige Nummer, die gleichzeitig der Primärschlüssel ist, einen Namen, der diesen Standard beschreibt und zu jedem Attribut die Informationen, ob der Standard nach diesem Attribut unterteilt ist und, wenn ja in welcher Genauigkeit. Der Name des Standards sollte zwar eindeutig gewählt sein, er eignet sich aber auf Grund seiner möglichen Länge nicht als Primärschlüssel.

Relation Externe_Fallzahlen_Metainfos

Attribut	Typ	Ausprägungen / Kommentare
<u>fallzahl_metainfo_id</u>	number(4)	Eindeutiger Schlüssel.
stdbev_name	char(40)	Name des externen Standards.
bezugsraum	char(6)	Teil des Schlüssels der Relation Standardbevolkerungen.
<u>faelle_inzidenz</u>	number(1)	Inzidenzzahlen vorhanden: 0 Nein / 1 Ja
<u>faelle_mortalitaet</u>	number(1)	Mortalitätszahlen vorhanden: 0 Nein / 1 Ja
<u>dim_geschlecht</u>	number(1)	Dimension vorhanden: 0 Nein / 1 Ja
<u>dim_diagnose</u>	number(1)	Dimension vorhanden: 0 Nein / 1 Ja
<u>icd_version</u>	number(2)	Fremdschlüssel aus K_ICD_Version.
<u>icd_stelligkeit</u>	number(1)	Stelligkeit der zugehörigen ICD-Kodes.
<u>dim_diagnose_anlass</u>	number(1)	Dimension vorhanden: 0 Nein / 1 Ja
<u>dim_diginitaet</u>	number(1)	Dimension vorhanden: 0 Nein / 1 Ja
<u>dim_histologie_d</u>	number(1)	Dimension vorhanden: 0 Nein / 1 Ja
<u>histologie_version</u>	char(2)	Fremdschlüssel aus K_Histlogie_Version.
<u>dim_diagnosesicherung</u>	number(1)	Dimension vorhanden: 0 Nein / 1 Ja
<u>dim_ausbreitung</u>	number(1)	Dimension vorhanden: 0 Nein / 1 Ja
<u>dim_tumorfolge</u>	number(1)	Dimension vorhanden: 0 Nein / 1 Ja

Entsprechend der Daten aus der Relation Externe_Fallzahlen_Metainfos werden in der Relation Externe_Fallzahlen die Fallzahlen zu einem externen Standard abgelegt. Wenn der externe Standard nach einem Attribut nicht unterteilt ist, d.h. diese Dimension nicht existiert, dann ist das entsprechende Attribut NULL. Ansonsten muß der Wertebereich der Ausprägungen dieses Attributs in der Relation Externe_Fallzahlen zu dem externen Standard bekannt sein. Dies wird dadurch gewährleistet, daß zu jedem Attribut eine Relation existiert, die zu einem externen Standard genau den Wertebereich des Attributs speichert. Die Namen dieser Relationen entsprechen

dem Namen des Attribut mit dem Zusatz „*Extern_*“. Das Attribut *Fallart* entscheidet, ob es sich um eine Mortalitätsfallzahl oder um eine Inzidenzfallzahl handelt

Das Attribut *Alter* muß immer genau den Wertebereich haben, den die zugehörige Standardbevölkerung aus der Relation Standardbevoelkerungen hat. Eine andere Alterseinteilung als die Einteilung, die bei den Standardbevölkerungsdaten benutzt wird, ist nicht zulässig.

Tupel, bei denen das Attribut *Fallzahl* 0 ist, werden in der Relation *Externe_Fallzahlen* nicht gespeichert. Aus diesem Grund enthält diese Relation meist weniger Tupel pro externem Standard als das Produkt der Mächtigkeiten der Wertebereiche der Dimensionen des Standards vermuten ließe. Die Attribute der Dimensionen, nach denen nicht unterschieden wird (d.h. deren Dimensionsattribut in *Externe_Fallzahlen_Metainfos* 0 ist), haben den Wert NULL.

Relation *Externe_Fallzahlen*

Attribut	Typ	Ausprägungen / Kommentare
<u>fallzahl_metainfo_id</u>	number(4)	Schlüssel aus <i>Externe_Fallzahlen_Metainfos</i> .
fallart	number(1)	1 Inzidenz / 2 Mortalität
geschlecht	number(1)	Teil des Schlüssels von der Relation <i>Standardbevoelkerungen</i> .
altermin	number(3)	Sterbe bzw. Diagnosealter abhängig von dem Attribut <i>fall_art</i> . Teil des Schlüssels von der Relation <i>Standardbevoelkerungen</i> .
diagnose	char(6)	< s. <i>Erweiterte_Tumordaten</i> >
diagnose_anlass	number(3)	< s. <i>Erweiterte_Tumordaten</i> >
diginität	number(1)	< s. <i>Erweiterte_Tumordaten</i> >
histologie_d	number(5)	< s. <i>Erweiterte_Tumordaten</i> >
diagnosesicherung	number(1)	< s. <i>Erweiterte_Tumordaten</i> >
ausbreitung	number(1)	< s. <i>Erweiterte_Tumordaten</i> >
tumorfolge	number(1)	< s. <i>Erweiterte_Tumordaten</i> >
fallzahl	number(10)	Fallzahl in der spezifizierten Gruppe.

Der Primärschlüssel würde normalerweise aus allen Attributen bis auf das Attribut *Fallzahl* bestehen, jedoch sind bei den Dimensionsattributen NULL-Werte erlaubt und somit ist dieser Schlüssel nicht erlaubt, obwohl die Werte dieser Attribute ein Tupel eindeutig identifizieren.

Extern_Geschlecht

Attribut	Typ	Ausprägungen / Kommentare
<u>fallzahl_metainfo_id</u>	number(4)	Fremdschlüssel aus <i>Externe_Fallzahlen_Metainfos</i> .
geschlecht	number(1)	< s. <i>Erweiterte_Tumordaten</i> >

Extern_Diagnose

Attribut	Typ	Ausprägungen / Kommentare
<u>fallzahl_metainfo_id</u>	number(4)	Fremdschlüssel aus <i>Externe_Fallzahlen_Metainfos</i> .
diagnose	char(6)	< s. <i>K_ICD</i> >

Extern_Diagnose_anlass

Attribut	Typ	Ausprägungen / Kommentare
<u>fallzahl_metainfo_id</u>	number(4)	Fremdschlüssel aus <i>Externe_Fallzahlen_Metainfos</i> .

<u>diagnoseanlass</u>	number(3)	< s. Erweiterte_Tumordaten >
-----------------------	-----------	------------------------------

Extern_Histologie_d

Attribut	Typ	Ausprägungen / Kommentare
<u>fallzahl_metainfo_id</u>	number(4)	Fremdschlüssel aus Externe_Fallzahlen_Metainfos.
<u>histologie_d</u>	number(5)	< s. Erweiterte_Tumordaten >

Extern_Dignitaet

Attribut	Typ	Ausprägungen / Kommentare
<u>fallzahl_metainfo_id</u>	number(4)	Fremdschlüssel aus Externe_Fallzahlen_Metainfos.
<u>diginitaet</u>	number(1)	< s. Erweiterte_Tumordaten >

Extern_Diagnosesicherung

Attribut	Typ	Ausprägungen / Kommentare
<u>fallzahl_metainfo_id</u>	number(4)	Fremdschlüssel aus Externe_Fallzahlen_Metainfos.
<u>diagnosesicherung</u>	number(1)	< s. Erweiterte_Tumordaten >

Extern_Ausbreitung

Attribut	Typ	Ausprägungen / Kommentare
<u>fallzahl_metainfo_id</u>	number(4)	Fremdschlüssel aus Externe_Fallzahlen_Metainfos.
<u>ausbreitung</u>	number(1)	< s. Erweiterte_Tumordaten >

Extern_Tumorfolge

Attribut	Typ	Ausprägungen / Kommentare
<u>fallzahl_metainfo_id</u>	number(4)	Fremdschlüssel aus Externe_Fallzahlen_Metainfos.
<u>tumorfolge</u>	number(1)	< s. Erweiterte_Tumordaten >

4.3 Indizes

Neben Primärschlüsselindizes sind auf folgenden Tabellen über die jeweils genannten Attributkombinationen Indizes definiert:

- Einwohner: Jahr, Gebiet
- Objektteile: ObjNr, ObjTeilNr sowie ObjNr
- Segmente: SegmentNr, PunktNr sowie SegmentNr
- Polygone: PolygonNr
- Lochzuordnung: ObjNr, ObjTeilNr, Loch sowie ObjNr
- Untersuchungsgebiete: Name
- Teilgebiete: Hauptgebiet

Anhang A: Kodierungen

A.1 Die Internationale Klassifikation der Krankheiten (ICD)

Die ICD (International Classification of Diseases) ist die wichtigste, weltweit anerkannte Diagnosenklassifikation in der Medizin [LGH95]. Sie definiert eine systematische Übersicht, in welcher Krankheitsbilder nach feststehenden Kriterien Krankheitsgruppen zugeordnet werden, und stellt damit eine statistische Klassifikation dar (im Unterschied zu einer Nomenklatur, die detailliert alle pathologischen Zustände getrennt aufführt). Die Grundlagen dieser Klassifikation wurden 1893 mit dem allgemeinen Todesursachenverzeichnis gelegt. Etwa alle 10 Jahre erscheinen verbesserte bzw. überarbeitete Versionen der ICD. Seit ihrer 6. Revision (1948) wird sie von der Weltgesundheitsorganisation (WHO) herausgegeben.

Das Klassifikationsschema basiert in den bisherigen Auflagen im wesentlichen auf der Ätiologie (den Ursachen) der Erkrankungen (im Gegensatz zur Anatomie, zur Natur des Krankheitsprozesses oder zur Betrachtung der eine Behandlung erfordernden Manifestationen).

A.1.1 9. Revision der ICD

Die 9. Revision der ICD [DIM93] wurde 1976 von der WHO verabschiedet und hatte noch bis Ende 1997 Gültigkeit und wird im ambulanten Bereich auch noch 1998 in Deutschland (vgl. Abschnitt A.1.2) eingesetzt. Sie verwendet numerische, bis zu 4-stellige Codes der Form xxx.x³². Als zusätzliche Klassifikation wird ein analog aufgebauter E-Kode angeboten, durch dessen Angabe die Ursache einer Schädigung (insb. bei Unfällen und Vergiftungen) genauer spezifiziert werden kann. Für einige Erkrankungen sind sowohl Einordnungen nach der Ätiologie als auch nach Manifestationen in bestimmten Organen möglich. Die entsprechenden ICD-Kodes werden durch ein nachgestelltes „†“ bzw. „*“ gekennzeichnet.

Die Systematik ist in 17 *Krankheitsklassen* unterteilt, die mit römischen Zahlen durchnummeriert sind. Die bösartigen Neubildungen stehen in der Krankheitsklasse „II. Neubildungen“. Die Krankheitsklassen sind jeweils in (bis zu 24) *Krankheitsgruppen*, z.B. Gruppe „150 – 159 Bösartige Neubildungen der Verdauungsorgane und des Bauchfells“, und die Krankheitsgruppen wiederum in (bis zu 20) dreistellige *Krankheitskategorien* unterteilt, z.B. „151 Bösartige Neubildung des Magens“, untergliedert. Die Krankheitskategorien können weiter in (bis zu 10) *Krankheitssubkategorien* spezifiziert werden, so beschreibt z.B. „151.1 Pylorus“ eine bösartige Neubildung des Magenausganges. Im allgemeinen werden hier die Endung „8“ für eine Überlappung mehrerer Bereiche, und die Endung „9“ für eine fehlende nähere Bezeichnung verwendet. Für spezielle Bereiche sind (bis zu 10) Unterteilungen in einer zusätzlichen 5. Stelle je Krankheitskategorie möglich, jedoch sind 5-stellige ICD-Angaben nicht üblich und in der vom BMG herausgegebenen Fassung gar nicht enthalten.

Die Krankheitsklasse „II. Neubildungen“ ist in folgende Krankheitsgruppen untergliedert:

³² Gelegentlich auftretende zweite Nachkommastellen sind offiziell nicht von Bedeutung.

- 140 – 149 Bösartige Neubildungen der Lippe, der Mundhöhle und des Rachens.
- 150 – 159 Bösartige Neubildungen der Verdauungsorgane und des Bauchfells.
- 160 – 165 Bösartige Neubildungen der Atmungs- und intrathorakalen Organe.
- 170 – 175 Bösartige Neubildungen der Knochen, des Bindegewebes, der Haut und der Brustdrüse.
- 179 – 189 Bösartige Neubildungen der Harn- und Geschlechtsorgane.
- 190 – 199 Bösartige Neubildungen sonstiger und nicht näher bezeichneter Sitze.
- 200 – 208 Bösartige Neubildungen des lymphatischen und hämatopoetischen Gewebes.
- 210 – 229 Gutartige Neubildungen.
- 230 – 234 Karzinoma in situ.
- 235 – 238 Neubildungen unsicheren Verhaltens.
- 239 Neubildungen unbekanntes Charakters.

Die ICD-Kodes 140 – 199 lassen sich hierbei zu folgenden Obergruppen zusammenfassen:

- 140 – 195 Bösartige Neubildungen, die vermutlich primär oder als solche bezeichnet sind mit näher bez. Sitz, ausgenommen lymphatisches und hämatopoetisches Gewebe. (Die Unterteilung dieser Obergruppe erfolgt – wie auch in den Gruppen 210-229, 230-234 und 235-238 – nach der Lokalisation des Ursprungssitzes des Tumors.)
- 196 – 198 Bösartige Neubildungen, die vermutlich sekundär oder als solche bezeichnet sind mit näher bez. Sitz.
- 199 Bösartige Neubildungen ohne nähere Angabe des Sitzes.

A.1.2 10. Revision der ICD

Bei der 1993 in Kraft getretenen und ab dem 1.1.1998 auch in Deutschland offiziell für die amtliche Todesursachenstatistik eingeführten 10. Revision der ICD [DIM95] wurde das traditionelle einachsige Klassifikationsmodell beibehalten. Allerdings wurde das bisherige numerische Verschlüsselungssystem durch ein alphanumerisches (1 Buchstabe, gefolgt von 3 Ziffern, in der Form Xxx.x³³) ersetzt. Dadurch wurde ein breiterer Verschlüsselungsrahmen zur Verfügung gestellt und mehr Raum für künftige Revisionen gelassen, ohne daß es dabei wie früher zu Verschiebungen im Numerierungssystem kommen muß. Insgesamt gibt es jetzt 21 Kapitel. Generell sind die Subkategorien „7“ für „Sonstige“, „8“ für „mehrere Teilbereiche überlappend“ und „9“ für „nicht näher bezeichnet“ vorgesehen.

Das Kapitel Neubildungen gliedert sich (zunächst nach dem biologischen bzw. biotischen Verhalten, dann weiter hauptsächlich nach der Lokalisation) in folgende Gruppen:

- C00 – C75 Bösartige Neubildungen an genau bezeichneten Lokalisationen, als primär festgestellt o. vermutet, exkl. lymphatisches, blutbildendes o. verwandtes Gewebe.
- C00-C14 Lippe, Mundhöhle und Pharynx.

³³ Wie schon in der ICD-9 ist auch hier eine weitere, zweite Nachkommastelle unüblich.

- C15-C26 Verdauungsorgane.
- C30-C39 Atmungsorgane und sonstige intrathorakale Organe.
- C40-C41 Knochen und Gelenkknorpel.
- C43-C44 Haut.
- C45-C49 Mesotheliales Gewebe und Weichteilgewebe.
- C50 Brustdrüse.
- C51-C58 Weibliche Genitalorgane.
- C60-C63 Männliche Genitalorgane.
- C64-C66 Harnorgane.
- C69-C72 Auge, Gehirn und sonstige Teile des Zentralnervensystems.
- C73-C75 Schilddrüse und sonstige endokrinen Drüsen.
- C76 – C80 Bösartige Neubildungen ungenau bezeichneter Lokalisationen, sekundärer und nicht näher bezeichneter Lokalisationen (insb. C80 für bösartige Neubildungen ohne Angabe der Lokalisation).
- C81 – C96 Bösartige Neubildungen des lymphatischen, blutbildenden und verwandten Gewebes, als primär festgestellt oder vermutet.
- C97 Bösartige Neubildungen als Primärtumoren an mehreren Lokalisationen
- D00 – D09 In-situ-Neubildungen.
- D10 – D36 Gutartige Neubildungen.
- D37 – D48 Neubildungen mit unsicherem oder unbekanntem Verhalten.

A.2 Die ICD für die Onkologie (ICD-O)

Zur speziellen Unterstützung der Kodierung von Tumorerkrankungen (gut- und bösartigen Neubildungen) wurde neben der ICD die ICD-O (International Classification of Diseases for Oncology) entwickelt, die 1976 in erster Auflage herauskam. Die ICD-O ist im Gegensatz zur ICD eine zwei- (drei-)achsige Klassifikation, die aus einem Topographie- und einem Morphologieteil besteht. Letzter zerfällt in die Angabe des histologischen Typs der Neubildung und den Malignitätsgrad (Verhalten, Charakter, Dignität), der hinter einem „/“ angefügt ist.

Der Topographieteil ist in weiten Teilen mit der ICD identisch (für die erste Auflage mit der ICD-9). Er hat jedoch für den Bereich *aller* Neubildungen die gleichen Schlüssel, welche die ICD nur für die *bösartigen* Neubildungen (Krebs) verwendet. (Dies ist jetzt aufgrund der zusätzlichen Klassifikation nach Morphologie und vor allem Dignität möglich.) Somit wird eine größere Spezifität nach dem Sitz auch für andere (gutartige, sekundäre (metastatische), in situ und ungewisse oder unbekannte) Neubildungen bzw. morphologische Typen (Neubildungen des lymphatischen, blutbildenden oder verwandten Gewebes) ermöglicht, für die die ICD gar keine oder keine so ausführliche topographische Klassifizierung enthält. Fünfstellige Codes wie in der ICD gibt es in der ICD-O nur für die deutsche Auflage (s. Abschnitt A.2.3).

Der Morphologiekode besteht aus einer vierstelligen Zahl zwischen 8000 und 9999 mit einem vorangestellten „M“. Dreistellige Morphologiekodes spezifizieren Gruppen der entsprechenden vierstelligen Schlüsselnummern. Gewöhnlich besitzt ein höherer gegenüber einem niedrigeren Morphologiekode die größere Spezifität und ist somit im Zweifelsfall bei der Kodierung vorzuziehen. (Diese Aussage hat für die 2. Auflage der ICD-O (s. Abschnitt A.2.2) nur noch eingeschränkte Gültigkeit.) Der Morphologiekode 8000 beschreibt nicht näher spezifizierte Tumoren.

Einige Neubildungen sind spezifisch für bestimmte Lokalisationen oder Gewebetypen. Dies definiert eine Vielzahl von Konsistenzregeln zwischen Morphologie und Topographie.

Der Malignitätsgrad kann folgende Ausprägungen annehmen:

- 0 benigne oder gutartig
- 1 unbestimmt, ob gut- oder bösartig; Grenzfälle
- 2 Karzinoma in situ
- 3 maligne oder bösartig, primärer Sitz
- 6 maligne oder bösartig, sekundärer Sitz; Metastase
- 9 maligne oder bösartig, unbestimmt, ob primärer oder sekundärer Sitz

Die folgende Tabelle zeigt die Gegenüberstellung des Schlüssels für die Morphologie nach der ICD-O mit den entsprechenden Gruppen der ICD-9 und ICD-10.

ICD-9	ICD-10	Dignität	
210-229	D10-36	0	Gutartiger Tumor
239	D37-48	1	Tumor unbekanntes Charakters
235-238	D37-48	1	Tumor unsicheren Verhaltens
230-234	D00-09	2	Karzinoma in situ
140-195 u. 200-208	C00-76 u. C81-97	3	Bösartiger Tumor mit Angabe / Unterstellung, primär zu sein
196-198	C77-79	6	Bösartiger Tumor mit Angabe / Unterstellung, sekundär (Metastase) zu sein
199	C80	3, 6 (9)	Bösartiger Tumor unbekannter Lokalisation (primär oder sekundär)

A.2.1 1. Auflage der ICD-O

Die 1. Auflage der ICD-O wurde 1976 von der WHO veröffentlicht [WHO76] und basiert bezüglich des Topographieteils auf der Krankheitsklasse „II. Neubildungen“ der ICD-9. Es werden hierbei nur die ICD-9-Kodes der bösartigen Neubildungen (140 – 199) verwendet – nicht berücksichtigt werden die Krankheitsgruppen 200-208, 210-229, 230-234, 235-238 und 239 (vgl. Abschnitt A.1.1).

A.2.2 2. Auflage der ICD-O

Die 1990 erschienene 2. Auflage der ICD-O [PVM90] basiert auf dem Kodierungssystem der ICD-10, die 1993 publiziert wurde. Ihr Topographieteil lehnt sich an das Kapitel II der ICD-10 an. Die Sachverhalte werden durch einen vierstelligen Kode repräsentiert, der von C00.0 bis C80.9 läuft. Ein Dezimalpunkt (.) zeigt die Unterteilung der dreistelligen Kategorien an.

Der alphabetische Index der ICD-10 weist unter dem Schlagwort „Neoplasma“ eine fünfspaltige Tabelle mit den Überschriften Maligne, Sekundär oder metastatisch, In situ, Benigne, Unsicherer bzw. unbekannter Charakter auf. Für die Tumoren der Lunge beispielsweise:

	Maligne	Sekundär o. metastatisch	In situ	Benigne	Unsicherer o. unbekannter Charakter
Lunge	C34.9	C78.0	D02.2	D14.3	D38.1

Bei der ICD-O gibt es nur einen Topographiekode für die nicht näher bezeichnete Lunge, nämlich C34.9. Der Malignitätsgrad ist Teil des mit M beginnenden Morphologiekodes und wechselt entsprechend der Natur des Tumors. Beispielsweise wird ein Plattenepithelkarzinom der Lunge als C34.9 – M8070/3 erfaßt, ein gutartiges Adenom der Lunge als C34.9 – M8140/0.

Die C00-C97-Kategorien der ICD-10 enthalten einige Rubriken, die entweder morphologisch definiert sind oder sekundäre bzw. metastatische Tumoren bezeichnen. Diese sind in der ICD-O weggelassen worden. Es handelt sich um folgende ICD-10-Kategorien:

- C43 Melanom der Haut
- C45 Mesotheliom
- C46 Kaposi-Sarkom
- C78 Sekundäre Neoplasmen des Respirations- und Verdauungssystems
- C79 Sekundäre bösartige Tumoren anderer Lokalisationen
- C81-C96 Maligne Neoplasmen der lymphoiden, hämatopoetischen und verwandter Gewebe
- C97 Bösartige Tumoren multipler (unabhängiger) Primärlokalisierung

Die Rubriken C81-C96 der ICD-10 beispielsweise erhalten (ebenso wie D45-D47) in der ICD-O einen spezifischen Morphologiekode (M959-998) mit der Dignität 3 (bzw. 1) sowie einen entsprechenden Topographiekode (i.a. C42 oder C77) aus dem Bereich C00-C80.

Die Rubrik C97 der ICD-10 entfällt, da in der ICD-O multiple (unabhängige) Primärtumoren separat zu verschlüsseln sind.

A.2.3 Deutsche Auflage der ICD-O (ICD-O-DA)

Zu den internationalen (englischen) Versionen der ICD-O existieren (neben wörtlichen deutschen Übersetzungen) jeweils auch deutsche Auflagen, die ICD-O-DA, die sich zwar auf die Originalversionen abstützen, aber doch geringfügig – u.a. auch durch den zeitlichen Verzug – von diesen unterscheiden. Die ICD-O-DA zerfällt jeweils in den Lokalisationsschlüssel, der dem Topogra-

phieiteil der ICD-O entspricht, und den Histologieschlüssel, der den Morphologie teil repräsentiert.

Die 2. und 3. Auflage des Lokalisationsschlüssels basieren auf der ICD-O-1 und unterscheiden sich nur in einigen Details voneinander [Wag88]. Entsprechend übersetzen die 4. und dazu bis auf Fehlerkorrekturen identische 5. Auflage die ICD-O-2 [Wag93]. Die 1. Auflage des Histologieschlüssels [Her78] entspricht der Morphologie aus der ICD-O-1; die Übertragung der 2. Auflage der ICD-O ist in [WW97] erschienen.

A.3 TNM

Das TNM-System behandelt in erster Linie die Klassifikation maligner Tumoren nach der klinisch und – falls möglich – histopathologisch bestimmten anatomischen Ausdehnung der Erkrankung. Es wurde in den Jahren 1943-52 von P. Denoix (Frankreich) entwickelt und seitdem mehrfach überarbeitet. Derzeit ist die 2. Revision der 4. Auflage aktuell [HSSW92]; im Herbst 1997 ist die 5. Auflage [WW97] erschienen. Das Klassifikationssystem

- ist in seinen grundlegenden Prinzipien ungeachtet der Behandlung auf alle anatomischen Bezirke anwendbar und
- läßt spätere Ergänzungen durch Informationen, die erst durch histopathologische Untersuchung oder chirurgische Eingriffe erhältlich sind, zu.

Das TNM-System zur Beschreibung der anatomischen Ausdehnung der Erkrankung beruht auf der Feststellung der 3 Komponenten

- T Ausdehnung des Primärtumors,
- N Fehlen oder Vorhandensein und Ausdehnung von regionären Lymphknotenmetastasen,
- M Fehlen oder Vorhandensein von Fernmetastasen.

Durch das Hinzufügen von Ziffern zu diesen 3 Komponenten wird das Ausmaß der malignen Erkrankung angezeigt (mit steigender Größe):

T0, T1, T2, T3, T4; N0, N1, N2, N3; M0, M1.

Für jede Lokalisation existieren spezielle Regeln zur Vornahme dieser Einteilung. Ein X anstelle der Ziffer steht jeweils für „nicht beurteilbar“. „Tis“ bezeichnet „Carcinoma in situ“, „Ta“ spezielle papilläre Harnblasentumoren. Bestehen Zweifel bei der Festlegung der TNM-Kategorien, so soll die niedrigere, d.h. weniger fortgeschrittene Kategorie gewählt werden. Im Grunde ist das System eine „Kurzschrift“ zur Beschreibung der Ausdehnung eines bestimmten malignen Tumors.

Für jede Lokalisation werden 2 Klassifikationen beschrieben:

- Die klinische Klassifikation (prätherapeutische klinische Klassifikation) wird als TNM oder cTNM bezeichnet. Sie basiert auf vor der Behandlung erhobenen Befunden. Solche ergeben sich aufgrund von klinischer Untersuchung, bildgebenden Verfahren, Endoskopie, Biopsie, chirurgischer Exploration und anderen relevanten Untersuchungen.
- Die pathologische Klassifikation (postoperative histopathologische Klassifikation) wird als pTNM bezeichnet. Bei dieser Klassifikation wird der vor der Behandlung festgestellte Befund ergänzt oder abgeändert durch Erkenntnisse, die beim chirurgischen Eingriff und

durch die pathologische Untersuchung gewonnen werden. Die pathologische Beurteilung des Primärtumors (pT) erfordert eine Resektion des Primärtumors oder Biopsien, die zur Bestimmung der höchsten pT-Kategorie adäquat sind. Die pathologische Beurteilung der regionären Lymphknoten (pN) erfordert die Entfernung von Lymphknoten in einem Ausmaß, das die Aussage über das Fehlen regionärer Lymphknotenmetastasen (pN0) verlässlich macht und andererseits zur Bestimmung der höchsten pN-Kategorie ausreicht. Die pathologische Feststellung von Fernmetastasen (pM) erfordert die mikroskopische Untersuchung.

Die T-, N- und M-Kategorien können jeweils (abhängig von der Art des Tumors) zu Stadien von 0 bis 4 gruppiert werden.

Im Fall multipler simultaner Tumoren in einem Organ sollte der Tumor mit der höchsten T-Kategorie klassifiziert werden und die Multiplizität oder die Anzahl der Tumoren in Klammern angegeben werden, z.B. T2(m) oder T2(5). Bei simultanen bilateralen Krebsen paariger Organe soll jeder Tumor für sich klassifiziert werden. Bei einigen Tumoren ist der Faktor Multiplizität ein Kriterium der T-Klassifikation.

Jedes T, N oder M kann ggf. für spezielle Zwecke noch in Untergruppen der Form T2a, T2b oder sogar weiter T2a1, ... unterteilt werden. Weiterhin ist in der M-Kategorie eine Spezifikation des Ortes der Fernmetastasen durch einen Kode aus 3 Buchstaben möglich.

Zur Kennzeichnung von speziellen Fällen in der TNM- oder pTNM-Klassifikation werden die Symbole y, r, a und m benutzt:

- Das a-Symbol kennzeichnet Fälle, bei denen die Klassifikation erst anlässlich einer Autopsie erfolgte.
- Das Suffix „(m)“ wird benutzt, um multiple Primärtumoren in einem anatomischen Bezirk anzuzeigen.
- Wenn die Klassifikation während oder nach initialer multimodaler Therapie erfolgt, werden die TNM- oder pTNM-Kategorien durch das Präfix „y“ gekennzeichnet (z.B. yT2N1M0 oder ypT2pN2pM0).
- Rezidivtumoren nach krankheitsfreiem Intervall werden durch das Präfix „r“ gekennzeichnet (z.B. rT2N0M0 oder rpT3pN1pMX).

Zusammen mit der TNM-Kodierung wird oft auch das histopathologische Grading (der Differenzierungsgrad) des Tumors angegeben:

- GX Differenzierungsgrad kann nicht bestimmt werden,
- G1 Gut differenziert,
- G2 Mäßig differenziert,
- G3 Schlecht differenziert,
- G4 Undifferenziert.

Weiterhin kann zur Angabe der von den verwendeten diagnostischen Methoden abhängigen Zuverlässigkeit der Klassifikation ein einstelliger C-Faktor (Certainty Factor, Diagnosesicherung) hinter die Kategorien T, N und M gesetzt werden (z.B. T3C2N2C1M0C2). Es steht

- C1 für Ergebnisse aufgrund von diagnostischen Standardmethoden, z.B. Inspektion, Palpation und Standard-Röntgenaufnahmen, intraluminale Endoskopie bei bestimmten Organen,

- C2 für Ergebnisse aufgrund spezieller diagnostischer Maßnahmen, z.B. bildgebende Verfahren: Röntgenaufnahmen in speziellen Projektionen, Schichtaufnahmen, Computertomographie (CT), Sonographie, Lymphographie, Angiographie; nuklearmedizinische Untersuchungen; Kernspintomographie (NMR); Endoskopie, Biopsie und Zytologie,
- C3 für Ergebnisse aufgrund chirurgischer Exploration einschließlich Biopsie und zytologischer Untersuchung,
- C4 für Ergebnisse nach definitiver chirurgischer Behandlung und pathologischer Untersuchung des Tumorsektats sowie
- C5 für Ergebnisse aufgrund einer Autopsie.

Die klinische TNM entspricht den Graden C1, C2 und C3, die pathologische Klassifikation im allgemeinen dem Grad C4.

Das Fehlen oder Vorhandensein von Residualtumor (Resttumor) nach Behandlung wird durch die R-Klassifikation beschrieben. Ihre Verwendung ist fakultativ:

- RX Vorhandensein von Residualtumor kann nicht beurteilt werden,
- R0 Kein Residualtumor,
- R1 Mikroskopischer Residualtumor,
- R2 Makroskopischer Residualtumor.

A.4 Berufskodierung

Prinzipiell kommen zur Erfassung und Kodierung von Berufen zwei Schlüssel in Frage³⁴:

- Der nationale Schlüssel „Klassifizierung der Berufe“ (KldB) wird vom Statistischen Bundesamt herausgegeben [SBu92]. Es ist ein im Hauptteil 3-stelliger Kode. Die neueste Ausgabe 1992 revidiert die ältere von 1970/75, und die aktuellen Verschlüsselungen sind auf Datenträger 6-stellig, in Papierform 4-stellig verfügbar.
- Der internationale Schlüssel „International Standard Classification of Occupation“ (ISCO) ist ein 5-stelliger Kode, „betreut“ von der International Labour Organisation (ILO) [ZUMA85]. Die aktuelle Ausgabe ist von 1988, die Vorgänger-Version von 1968. Dieser Schlüssel liegt als englisch-sprachige Version – auch auf Diskette – vor. Eine deutsche Übersetzung der Berufsbezeichnungen, nicht aber der Tätigkeitsbeschreibungen, gibt es inzwischen auf Papier. Darüber hinaus gibt es von der aktuellen ISCO-Version eine EU-Variante: ISCO-COM.

In Absprache mit der Arbeitsgemeinschaft bevölkerungsbezogener Krebsregister in Deutschland [Sch96] soll in Niedersachsen die jeweils aktuelle Fortschreibung der KldB durch die Bundesanstalt für Arbeit verwendet werden. Diese bildet eine Erweiterung bzw. Aktualisierung des Schlüssels des statistischen Bundesamtes. Der Schlüssel ist hierarchisch in 5 Ebenen aufgebaut:

- 6 Berufsbereiche, z.B. „III Fertigungsberufe“,

³⁴ Nach Recherchen am Bremer Institut für Präventionsforschung und Sozialmedizin (BIPS) unter der Leitung von Frau Dr. van der Linde.

- 33 Berufsabschnitte, z.B. „IIIg Schlosser, Mechaniker und zugeordnete Berufe“,
- 86 Berufsgruppen, z.B. „25 Schmiede“,
- 334 Berufsordnungen, z.B. „251 Stahlschmiede“, und
- 1993 Berufsklassen, z.B. „2511 Hammerschmiede“.

Berufsgruppen, -ordnungen und -klassen bilden eine Dezimalklassifikation. Jeder Berufsklasse sind wiederum mehrere synonyme oder spezialisierende Berufsbezeichnungen zugeordnet. Die Zugehörigkeit zu Gewerbebranchen wird leider nur sehr eingeschränkt berücksichtigt.

Im Datenschema des EKN wird die Kodierung nach Gruppen, Ordnungen und Klassen unter Heranziehung der Berufsbezeichnungen unterstützt.

A.5 Staatenkodierung

Die Staatenkodierung erfolgt nach dem Gebietsschlüssel des Statistischen Bundesamtes. Die Schlüssel

- 001 bis 016 repräsentieren hierin die Bundesländer Deutschlands (000 steht für Deutschland),
- 101 bis 199 bezeichnen Länder Europas,
- 201 bis 299 stehen für Afrika,
- 301 bis 399 für Amerika,
- 401 bis 499 für Asien und
- 501 bis 599 für Australien und Ozeanien.

995 steht für „von / nach See“, 996 für unbekanntes Ausland, 997 repräsentiert Staatenlose, 998 solche mit ungeklärter Staatsangehörigkeit, und 999 steht für „fehlende Angabe“.

x95 ($x \in \{1, 2, 3, 4, 5\}$) bezeichnet jeweils die britisch abhängigen Gebiete, x99 den übrigen Kontinent.

Der offizielle Schlüssel wurde ergänzt um die Codes 100, 200, 300, 400 und 500 jeweils für die gesamten Kontinente (ohne nähere Angabe).

A.6 Gemeindekennziffern und regionale Beobachtungseinheiten

Gemeindekennziffern stellen eine Kodierung der Verwaltungseinheiten Deutschlands dar. Eine Gemeindekennziffer ist eine 8-stellige Zahl, von der

- die ersten zwei Stellen das Bundesland (3 für Niedersachsen),
- die dritte Stelle den Regierungsbezirk,
- die vierte und fünfte Stelle den Landkreis bzw. die kreisfreie Stadt und

- die letzten drei Stellen die Gemeinde oder gemeindefreie Gebiete (Samtgemeinden haben an der sechsten Stelle stets eine 4)

spezifizieren. Gemeindekennziffern von Verwaltungseinheiten höherer Aggregierungsebene als Gemeinde enden jeweils auf entsprechend viele Nullen. Gemeindekennziffern zu Landkreisen, von denen die drei Nullen am Ende abgeschnitten sind, werden auch als Kreiskennziffern bezeichnet (Bezirkskennziffern und Landeskennziffern analog).

Unter Anlehnung an die Gemeindekennziffern sollen in Niedersachsen sogenannte „regionale Beobachtungseinheiten“ definiert werden, die einheitlich zwischen 5.000 und 30.000 Einwohner haben. Diese Einheiten dienen zum einen zur Erfassung von Pathologienmeldungen unter Gewährleistung einer faktischen Anonymisierung auch in Gemeinden mit sehr kleinen Einwohnerzahlen. Zum anderen bilden sie einen einheitlichen Raumbezug mit geringerer Variabilität für Monitoring und Berichterstattung.

Regionale Beobachtungseinheiten werden durch Zusammenfassung benachbarter Kleingemeinden bzw. durch Unterteilung von Großgemeinden oder Städten (in Orts- oder Stadtteile) gebildet. Hierbei wird nach Möglichkeit auf ähnliche Bevölkerungs- und Wohnstrukturen in den einzelnen Einheiten geachtet. Kreisgrenzen bleiben auf jeden Fall erhalten, d.h. nur Gemeinden innerhalb eines Landkreises können verschmolzen werden. Ein genauer Algorithmus bzw. die zugrundeliegenden Kriterien zur Definition der Beobachtungseinheiten werden derzeit noch erarbeitet.

A.7 Gauß-Krüger-Koordinaten

Die seit 1927 in Deutschland verwendeten Gauß-Krüger-Koordinaten dienen zur Beschreibung der Lage von Punkten auf der Erde.

Die Grundlage der Abbildung der Erdoberfläche in eine Ebene mittels der Gauß-Krüger-Koordinaten bildet die sogenannte *Gaußsche konforme Abbildung GkA*. Ein Längengrad wird durch diese jeweils als Bezugsmeridian ausgewählt und längentreu und gradlinig, d.h. winkeltreu (die Winkel bleiben bei dieser Abbildung erhalten) abgebildet. Alle anderen Punkte des Geoids werden bezüglich dieses Meridians abgebildet. Hierbei treten höhen- und entfernungsabhängige Längenverzerrungen auf. Die Ungenauigkeit wächst quadratisch mit der Entfernung zum Bezugsmeridian.

Um die Abweichung in Grenzen zu halten, wird die *GkA* in Teilabbildungen zerlegt. Jede dieser Teilabbildungen bildet nun nur noch einen Streifen längs eines bestimmten Längengrades in die Ebene ab. Jeder dieser Streifen hat seinen eigenen Bezugsmeridian, der jeweils längentreu abgebildet wird. Alle Punkte innerhalb eines Streifens werden dann mit Bezug auf ihren Bezugsmeridian abgebildet.

In Deutschland werden 3° breite Streifen mit den Längengraden 6°, 9°, 12° und 15° östlicher Länge als Bezugsmeridiane verwendet. Jeder der erwähnten Streifen liegt jeweils 1,5° rechts und links um diese Bezugsmeridiane herum. Durch diese Aufteilung der Abbildung auf 3°-Streifen wird eine obere Schranke für Entfernungen von Punkten zum jeweiligen Bezugsmeridian eingeführt und dadurch die Längenverzerrung begrenzt.

Um die Gauß-Krüger-Koordinaten eines bestimmten Punktes zu finden, berechnet man zuerst seinen Abstand (in Metern) vom Äquator und von seinem nächsten Bezugsmeridian. Der Abstand des Punktes vom Äquator wird mit x , der vom Bezugsmeridian mit y bezeichnet. Liegt der

Punkt westlich vom Bezugsmeridian, so wird y mit negativem Vorzeichen, östlich vom Bezugsmeridian mit positivem Vorzeichen versehen. x und y sind die *Gaußschen konformen Koordinaten* dieses Punktes.

Desweiteren muß die Kennzahl des Bezugsmeridians ermittelt werden: $K = L_M / 3^\circ$, mit L_M als Längengrad des Bezugsmeridians. Im Gauß-Krüger System werden Abszissen als *Hochwerte* H , Ordinaten als *Rechtswerte* R bezeichnet. H und R stehen mit den Gaußschen konformen x - und y -Koordinaten in folgender Beziehung:

$$H = x$$

$$R = K \cdot 10^6 + c + y$$

mit $c = 500.000$ m, damit auch westlich des Bezugsmeridians positive Rechtswerte gelten.

Hoch- und Rechtswert haben in Deutschland also jeweils sieben Ziffern. Sie können auch als reelle Zahlen angegeben werden, wenn die Genauigkeit der Koordinaten feiner als 1m sein soll.

Durch die Winkeltreue der Gauß-Krüger-Koordinaten kann mit diesen genau wie mit normalen Koordinaten in der Ebene gerechnet werden. Dies gilt jedoch nur, solange man Punkte in einem einzigen Meridianstreifen betrachtet. Die Koordinaten von Punkten benachbarter Streifen müssen zunächst in Koordinaten im Meridianstreifen des anderen Punktes umgewandelt werden.

Anhang B: Datenbank-Installation

Die Installation einer Registerstellendatenbank³⁵ besteht aus zwei Schritten. Erst wird das Schema der Datenbank installiert (siehe Abschnitt B.2), und danach werden die Tabellen mit den Geo-Daten und den Kodierungen geladen (siehe Abschnitt B.3). Sowohl die Schemainstallation als auch das Laden der Daten werden über jeweils ein Skript gestartet. Diese Skripte führen in Unterverzeichnissen weitere Skripte aus, die dann das Schema bzw. die Daten laden. Das Installationsverzeichnis der Registerstellendatenbank hat den Namen `DBInstall`. In diesem Verzeichnis finden sich zahlreiche Unterverzeichnisse sowie Skripte, deren Bedeutung im Abschnitt B.1 erklärt wird.

B.1 Verzeichnisstruktur

Das Verzeichnis `DBInstall` hat die folgenden Unterverzeichnisse:

- `Schema_Beschreibung` Installationsverzeichnis für das Schema
- `Data` Skripte für den Benutzer zum Laden von Daten
 - `README` readme-Dateien zu den Lade-Skripten
 - `Einwohnerdaten.Niedersachsen`
 - `...` Hilfsdateien für das Laden der Einwohnerdaten
- `Geo_Daten` Geo-Daten der Untersuchungsgebiete
 - `Niedersachsen` Geo-Daten für Niedersachsen
 - `Geometrie_Daten` Geometrien der Geo-Objekte
 - `Region_Zusaetze` Zusätzliche Informationen zu den Regionen Niedersachsens
 - `Strassen` ...
 - `Kraftwerke` ...
 - `...` ...
 - `Einwohner` Einwohnerzahlen von Niedersachsen
 - `Flaechennutzung` Flächennutzung in Niedersachsen
 - `ReBes` Definition der regionalen Beobachtungseinheiten
 - `WeserEms` Geo-Daten für das Gebiet Weser-Ems
 - `...` Die Struktur ist identisch zu den niedersächsischen Daten
 - `NULLValues` NULL-Werte für die Geo-Daten
- `Indizes` Definition der Indizes auf dem Schema
- `Kodierungen` Kodierungstabellen
 - `Diagnosen` Kodierungen der Diagnosen
 - `Histologien` ...
 - `Lokalisation` ...

³⁵ Die Installation der Vertrauensstellen-Datenbank läuft nach ähnlichem Prinzip – hierauf soll in diesem Bericht nicht näher eingegangen werden.

• ICD_Codes	...
• Versionen	Aktuell benutzte Versionen der Diagnosen
• Berufe	Kodierungen der Berufe
• Staaten	Kodierungen der Staaten
• StdBevDaten	Daten zu den Standardbevölkerungen
• Einwohner	Einwohnerzahlen der Standardbevölkerungen
• Fallzahlen	Fallzahlen der Standardbevölkerungen
• Saarland	Fallzahlen für die Standardbevölkerung des Saarlands
• Log_Files	Log-Files aus vorhergehenden Installationen
• Skripte	Hilfsskripte
• Erweiterte_TD_fuellen	Übertragen der Meldungen in materialisierten View

Die Aufgaben dieser Unterverzeichnisse sollen in den folgenden Abschnitten genauer erklärt werden.

B.2 Schemainstallation

Die Schemainstallation umfaßt die folgenden Arbeitsschritte:

1. Löschen der existierenden Relationen
2. Erstellen der Relationen
3. Definieren der Integritätsbedingungen
 - Primärschlüssel festlegen
 - Fremdschlüssel festlegen

Bei der Schemainstallation wird ein existierendes Schema **überschrieben**, die existierenden Relationen und deren Inhalte werden **gelöscht**. Bevor diese Skripte ausgeführt werden, muß man sich über diese Tatsache im klaren sein. Das **vollständige** Schema wird durch das Kommando

```
Create_Reg <DB-Benutzer> <DB-Password>
```

geladen. Um überprüfen zu können, ob die Installation fehlerfrei verlaufen ist, ist es sinnvoll, die Ausgabe des Skriptes in eine Datei umzuleiten, z. B.:

```
Create_Reg peggy sillyPasswd > myDBInstallation.log
```

Diese Datei enthält dann die Ausgaben der von diesem Skript aufgerufenen Kommandos. Diese Log-Datei sollte in das Verzeichnis Log_Files kopiert werden, um auch später noch Zugriff darauf zu haben. In einer Log-Datei kann man beispielsweise mit „grep ORA myDBInstallation.log“ nach Oracle-Fehlern suchen. Die Zuordnung einer Fehlermeldung zu einem Arbeitsschritt ist jedoch nur mit Kenntnis des internen Aufbaus der Skripte möglich. In Abschnitt B.2.1 wird erklärt wie die Wertebereiche der Attribute definiert sind und in den Abschnitten B.2.2 bis B.2.5 wird die Schemainstallation der Teilschemata der Kodierungstabellen, Geo-Daten, Fall- und Meldungsdaten sowie der Bevölkerungsdaten beschrieben.

B.2.1 Wertebereiche der Attribute

Alle Skripte zur Schemainstallation befinden sich in dem Verzeichnis `DBInstall/-Schema_Beschreibung`. Die Datei `domaenen` enthält C-Präprozessordefinitionen, die für jeden Wertebereich eines Attributs den entsprechenden SQL-Datentyp definiert. Bei Aufzählungstypen ist der Wertebereich des Attributs durch einen Check-Constraint eingeschränkt.

Das Schema der Registerstelle ist in mehrere Teilschemata unterteilt. Diese Teilschemata fassen alle Relationen zusammen, die zu einem thematisch abgeschlossenen Bereich gehören. Für jedes SQL-Skript zum Erstellen eines solchen Teilschemas wird eine Datei erstellt, die statt des SQL-Datentyps zur Attributdefinition das entsprechende C-Präprozessormakro benutzt. Aus dieser Datei kann der C-Präprozessor ein gültiges SQL-Skript erzeugen. Mit diesem Verfahren erreicht man, daß alle Attribute, die durch Konstruktion identische Wertebereiche haben, auch wirklich identische Definitionen in SQL haben. Gerade wenn ein Schema über längere Zeit gepflegt werden soll, kann es leicht vorkommen, daß per Definition identische Wertebereiche dennoch von einander abweichen, weil Änderungen nicht an allen notwendigen Stellen gleichermaßen durchgeführt wurden.

Beispiel für eine C-Präprozessordefinition:

```
#define DOM_raucher_status NUMBER(1) NOT NULL \
        CHECK ( raucher_status IN ( 1 , 2 , 3 , 9 ) )
```

Beispiel einer Quelldatei für ein SQL-Skript:

```
CREATE TABLE standardmeldungen(...,
        raucher_status DOM_raucher_status,...)

CREATE TABLE tumoren(...,
        raucher_status DOM_raucher_status,... )
```

Wenn sich ein Wertebereich eines Attributs geändert hat, müssen alle SQL-Skripte, die diese Definition benutzen, neu erzeugt werden. Das Kommando `Install_new_Domains` löscht alle alten Übersetzungsergebnisse und erzeugt dann die SQL-Skripte mit den Definitionen der Datei `domaenen` neu. Das Skript `Install_new_Domains` stellt ebenfalls sicher, daß die Wertebereiche der Attribute der Registerstelle mit denen der Vertrauensstelle übereinstimmen. Wichtig ist, daß die erzeugten SQL-Skripte (Endung `.sql`) nicht per Hand verändert werden dürfen, da diese bei jedem Aufruf von `Install_new_Domains` überschrieben werden. Alle Änderungen müssen in den zugehörigen C-Präprozessordateien (Endung `.cpp`) vorgenommen werden.

B.2.2 Schemainstallation der Kodierungstabellen

Bei der Installation des Schemas der Kodierungstabellen werden die folgenden Skripte benutzt:

- Die Kodierungstabellen können mit dem Skript `kdrops.sql` gelöscht werden.
- Mit dem Skript `k1.sql` werden die Skripte `k1.tab` und `k1.con` nacheinander gestartet.
- Das Skript `k1.tab` zum Erzeugen des Schemas der Kodierungstabellen wird mit dem Präprozessor aus der Datei `k1.tab.cpp` erzeugt. Dieses SQL-Skript erzeugt alle Relationen für die Kodierungen (vgl. Abschnitt 1.2.5).

- Das Skript `k1.con` ergänzt bei den Kodierungstabellen die Informationen über Fremd- und Primärschlüssel. Da dieses Skript keine Wertebereiche mehr kennen muß, braucht es nicht vom Präprozessor übersetzt werden.

Von dem Skript `Create_Reg` werden die existierenden Kodierungstabellen durch das Skript `kdrop.sql` gelöscht, und danach werden diese mit dem aktuellen Schema durch das Skript `k1.sql` neu erstellt.

Es gibt bisher keine Fremdschlüsselbeziehungen zwischen den Kodierungstabellen und den Tabellen der Registerstelle.³⁶ Wenn auf eine Ausprägung einer Kodierungstabelle referenziert wird, so umfaßt der Wertebereich dieses Attributs (eingeschränkt durch den Check-Constraint) genau den Wertebereich der Ausprägung in der Kodierungstabelle. Die Menge der Ausprägungen einer Kodierungstabelle muß also immer mit der Menge der erlaubten Werte im entsprechenden Check-Constraint übereinstimmen.

B.2.3 Schemainstallation der Geo-Daten

Bei der Installation des Geo-Daten-Schemas ist die Installation des Schemas für die Bevölkerungsdaten nicht enthalten, diese wird jedoch in Abschnitt B.2.5 beschrieben. Bei der Installation des Schemas der Geo-Daten werden die folgenden Skripte benutzt:

- Die Relationen mit den Geo-Daten werden mit dem Skript `gdrop.sql` gelöscht.
- Mit dem Skript `gcreate.sql` werden die Skripte `geo1.tab` und `geo1.con` nacheinander ausgeführt.
- Das Skript `geo1.tab` erzeugt das Schema der Geo-Daten (vgl. Kapitel 4).
- Das Skript `geo1.con` fügt die Fremd- und Primärschlüsselbedingungen hinzu.

Die existierenden Geo-Daten werden durch das Skript `gdrop.sql` gelöscht und danach mit dem aktuellen Schema durch das Skript `gcreate.sql` neu erstellt. Auf die in Abschnitt B.2.1 vorgestellte Domänen-Definition wird hier nicht zugegriffen.

B.2.4 Schemainstallation der Fall- und Meldungsdaten

Bei der Installation des Schemas der Fall- und Meldungsdaten der Registerstelle werden die folgenden Skripte benutzt:

- Die bestehenden Relationen werden mit dem Skript `rdrop.sql` gelöscht.
- Das Skript `rs.sql` zum Erzeugen der Relationen und der dazugehörigen Fremd- und Primärschlüsselbedingungen der Registerstelle wird mit dem Präprozessor aus dem Skript `rs.cpp` erzeugt.
- Die Datei `ems_only.sql` definiert weitere Integritätsbedingungen für die Relation `Erweiterte_Tumordaten`, die aber nur dann angewandt werden dürfen, wenn eine vollständige Registerstellendatenbank vorliegt. Diese Datei darf nicht ausgeführt werden, wenn die Relation `Erweiterte_Tumordaten` mit Daten gefüllt ist, zu denen keine Meldun-

³⁶ Eine Überarbeitung des Schemas in diese Richtung ist geplant.

gen existieren. Da bisher der Rechner der Registerstelle mit Namen „ems“ als einziger Meldung speichert, wird in dem Skript `Create_Reg` der Rechnername abgefragt und davon abhängig dieses Skript ausgeführt oder nicht.

Das Skript `Create_Reg` löscht die Relationen der Registerstelle mit dem Skript `rdrop.sql` und erstellt diese mit dem aktuellen Schema durch das Skript `rs.sql` und ggf. `ems_only.sql` neu.

B.2.5 Schemainstallation für die Bevölkerungsdaten

Bei der Installation des Schemas für die Bevölkerungsdaten des Studiengebiets und der externen Standardbevölkerungen mit und ohne Fallzahlen werden die folgenden Skripte benutzt:

- Die bestehenden Relationen werden mit dem Skript `adrop.sql` gelöscht.
- Das Skript `add_data.sql` zum Erzeugen der Relationen der Registerstelle wird mit dem Präprozessor aus dem Skript `add_data.cpp` erzeugt.

Das Skript `Create_Reg` löscht die Relationen der Registerstelle mit dem Skript `adrop.sql` und erstellt diese mit dem aktuellen Schema durch das Skript `add_data.sql` neu.

B.3 Laden der Daten

Die Daten, die sich nicht aus den eingehenden Meldungen ergeben, können durch Skripte geladen werden. Zu diesen Daten gehören

- Geo-Daten mit den Grenzzügen der Untersuchungsgebiete,
- Geo-Daten mit Gefahrenpunkten,
- Einwohnerdaten in den Untersuchungsgebieten,
- Standardbevölkerungen,
- externe Fallzahlen,
- Kodierungstabellen für Diagnose, Histologie, Beruf, usw.

Alle diese Daten sind im laufenden Betrieb kaum Änderungen unterworfen und sollen auf unterschiedlichen Installationen einer Registerstelle identisch sein. Daher wird davon ausgegangen, daß Änderungen der Daten nur in Loader-Skripten und nicht in der Datenbank durch Updates vollzogen werden. Wenn diese Daten in der Datenbank verändert würden, könnte nicht mehr garantiert werden, daß bei einer anderen Installation diese Änderungen nachvollzogen werden könnten.

Die Daten können entweder in ein leeres Datenschema vollständig geladen werden oder es können Teile der Daten einer Datenbank durch andere Daten ersetzt werden. Die meisten Daten werden aus Textdateien mit Hilfe des SQL-Loaders geladen. Datensätze, die gegen die Integritätsbedingungen (z.B. Primärschlüsselbedingung oder Check-Constraint) der zu ladenden Relation verstoßen, werden in eine Datei mit der Endung `.bad` geschrieben. Es ist daher sinnvoll, mit dem Kommando „`find . -name \"*.bad\"`“ nach dem Laden von Daten nach diesen Dateien zuzusehen. Wenn solche Dateien existieren, enthalten die Textdateien fehlerhafte Daten. Vor

jedem Laden sollten die alten `.bad` Dateien gelöscht, um diese nicht mit den neuen Dateien zu verwechseln.

B.3.1 Vollständiges Laden der Daten

Die Daten können mit dem Kommando

```
Fill_Tables <DB-Benutzer> <DB-Password> <Geo-Region>
```

vollständig geladen werden. Der Parameter `Geo-Region` bestimmt, für welchen Bereich die Geo-Daten mit den Einwohnerzahlen geladen werden sollen. Die derzeit möglichen Werte dieses Parameters sind:

- „Niedersachsen“ Geo-Daten für Niedersachsen
- „WeserEms“ Geo-Daten für Weser-Ems
- „all“ Geo-Daten für Niedersachsen und Weser-Ems

Die Ausgabe des Skripts `Fill_Tables` kann in eine Datei umgeleitet werden, die in das Verzeichnis `Log_Files` kopiert werden sollte, damit eventuelle Fehler nachvollzogen werden können.

Dieses Skript und alle von diesem Skript aufgerufenen Skripte zum Laden der Daten gehen davon aus, daß die Relationen leer sind. Ein doppelter Aufruf eines Loader-Skripts kann aus diesem Grund zu unerwünschten Ergebnissen führen. Wenn eine Datenbank bereits Daten enthält, so darf das Skript `Fill_Tables` also nicht aufgerufen werden. Aus diesem Grund existieren in dem Verzeichnis `data` einige Skripte, mit denen gezielt bestimmte Daten in eine existierende Datenbank nachgeladen werden können.

Fast alle Daten werden mit Hilfe des SQL-Loaders `sqlldr` in die Datenbank importiert, da dieses Verfahren bei großen Datenmengen eine erheblich bessere Performance und eine wesentliche höhere Flexibilität als die Insert-Anweisung von SQL bietet. Der SQL-Loader bekommt über Steuerdateien mit der Endung `.ctl` Informationen über den Aufbau der als Textdatei vorliegenden Daten. Der Aufbau dieser Steuerdateien ist in der einschlägigen Oracle-Literatur beschrieben.

B.3.1.1 Laden der Geo-Daten

Das Skript `load_data` in dem Verzeichnis `DBInstall/Geo_Daten` benötigt die gleichen Parameter wie das Skript `Fill_Tables` und lädt die Geo-Daten (vgl. Kapitel 4) für das angegebene Gebiet aus diversen Unterverzeichnissen.

Geometrien

Die Geometrie-Daten (vgl. Abschnitt 4.2.2) werden aus dem Verzeichnis `Geometrie_Daten` mit dem Skript `Nds_loader` geladen. Über die Steuerdateien werden hier (in der angegebenen Reihenfolge) die Daten der Punkte, Geo-Objekte, Objektteile, Segmente, Polygone, Darstellungsebenen, Untersuchungsgebiete, Gebiete, Teilgebiete und Ebenenzuordnungen geladen.

Nach dem Laden der Geometrien werden nachträglich Korrekturen mit dem Skript `update.sql` vorgenommen. Regionen, die bisher in keinem Landkreis enthalten waren, werden

einem Landkreis zugeordnet, und fehlerhafte Objektteile werden gelöscht. Da die Geo-Daten teilweise unvollständig sind, müssen die fehlerhaften bzw. fehlenden Mittelpunkts-Koordinaten einiger Gebiete durch das Skript `mittelpunkte.sql` durch die korrekten Koordinaten ersetzt.

Einwohner

Die Relation `Einwohnermetainfos` (vgl. Abschnitt 4.2.4) wird durch das Skript `loader` geladen und braucht nach dem Laden nicht weiter verändert werden.

Die Einwohnerdaten im Verzeichnis `Einwohner` werden von dem Skript `loader` für jedes in die Datenbank zu ladende Jahr einzeln durch das Skript `AddEinwohner` geladen. Das Skript `AddEinwohner` lädt die Einwohnerdaten zunächst in eine temporäre Relation, da diese noch korrigiert werden müssen und ansonsten gegen die Fremdschlüsselbedingungen der Relation `Einwohner` verstoßen würden. Die Einwohnerdaten der Regionen, zu denen keinen Geo-Daten existieren, werden gelöscht. Die Gemeindekennziffern in den Einwohnerdaten werden durch eine interne Nummerierung ersetzt. Nach diesen Korrekturen werden die Daten von der temporären Relation auf die Relation `Einwohner` übertragen.

Wegen der Fremdschlüsselbedingungen zwischen Geo- und Einwohnerdaten müssen die Einwohnerdaten immer nach den Geo-Daten geladen werden. Aus dem gleichen Grund muß auch die Relation `Einwohnermetainfos` vor der Relation `Einwohner` geladen werden, da die Altersgruppen dieser Relation die Grundlage für die Einwohnerzahlen in der Relation `Einwohner` bilden.

Flächennutzung

Informationen über die Flächennutzung der Regionen stehen in dem Unterverzeichnis `Flaechennutzung` und werden von dem Skript `loader` geladen (vgl. Abschnitt 4.2.4). Nach dem Laden müssen wie bei der Relation `Einwohner` erst die Gemeindekennziffern durch eine interne Nummerierung ersetzt und die Informationen zu in den Geo-Daten nicht verfügbaren Regionen entfernt werden.

Regionen-Zusätze

In dem Verzeichnis `Region_Zusaetze` sind Geo-Daten wie Straßen, Gewässer usw. enthalten. Diese Daten werden aber zur Zeit von keinem Werkzeug benutzt, und daher werden die Loader-Skripte der Daten an dieser Stelle nicht beschrieben.

NULL-Werte

Zur Repräsentation aller Patienten, deren Wohnort nicht bekannt ist, wird durch das Skript `nullregion` in dem Unterverzeichnis `NULLvalues` eine neue Region „Irgendwo“ angelegt. Damit bilden diese Patienten keinen Sonderfall mehr, sondern können über diese zusätzliche Region normal ausgewertet werden. Durch das Skript `nullregioninfo` werden dieser Region Flächennutzungsinformationen zugewiesen. Außerdem werden diese Informationen teilweise auch bei den Regionen ergänzt, die beim Laden der Geo-Daten korrigiert werden mußten.

Durch das Skript `nullaltersgruppe` wird eine neue Altersgruppe in die Relation `Einwohnermetainfo` eingetragen, die den Patienten zugeordnet wird, bei denen das Alter unbekannt ist.

Die genannten Skripte müssen, da sie auf verschiedenen anderen Relationen aufbauen, im Anschluß an das Laden aller anderen Geodaten ausgeführt werden.

Regionale Beobachtungseinheiten

In dem Verzeichnis `ReBes` sind Informationen enthalten, welche Gemeinden jeweils zu regionalen Beobachtungseinheiten (ReBes) aggregiert werden sollen. Um aus Gemeinden ReBes zu bilden, müssen die Geometrien, Einwohnerzahlen und Flächennutzungen für jede ReBe aus den Gemeinden aggregiert werden.

Die Geometrien liegen bereits in aggregierter Form als Loader-Dateien vor und werden durch das Skript `loader` geladen. Dieses Skript wird von dem Loader der Geo-Daten direkt aufgerufen.

Das Skript `AddReBesEinwohner` aggregiert die Einwohnerzahlen für die ReBes, und das Skript `AddReBesFlaechennutzung` aggregiert die Flächennutzungen für die ReBes. Diese beiden Skripte werden jeweils nach dem Laden der Einwohnerzahlen bzw. Flächennutzungen aufgerufen, um die fehlenden Zahlen für die ReBes zu ergänzen.

B.3.1.2 Erstellen von Indizes

Nachdem das Schema und die Geo-Daten installiert wurden, müssen die Indizes auf den Relationen festgelegt werden. Im Unterverzeichnis `Indizes` wird zu diesem Zweck von `Create_Reg` das Shell-Skript `Index_Creator` aufgerufen. Mit diesem Skript werden nacheinander die SQL-Skripte

- `geo_index` für die Indizes des Geo-Schema,
- `abgleich_index` für die Indizes zum Abgleich der Meldungen,
- `fall_index` für die Indizes zum Auswerten der Falldaten und
- `einwohner_index` für die Indizes der Einwohnerdaten

ausgeführt.

B.3.1.3 Laden der Standardbevölkerungen

Die Standardbevölkerungen werden aus dem Verzeichnis `StdBevDaten` mit dem Skript `StdBevDaten.loader` geladen. Die Altersgruppen der Standardbevölkerungen müssen identisch mit den Altersgruppen der Einwohnerdaten sein, da ansonsten diese Daten nicht miteinander verglichen werden können.

Die Bevölkerungszahlen finden sich in dem Unterverzeichnis `Einwohner`. In diesem Verzeichnis gibt es – entsprechend der gewählten Variante der Altersgruppierung – mehrere Versionen der Standardbevölkerungen, es wird aber nur die Version geladen, auf die der Link namens `standardbevoelkerung.data` zeigt. Das Skript `LoadStdPop.loader` lädt die Bevölkerungszahlen über diesen Link.

Die Fallzahlen sind in dem Unterverzeichnis `Fallzahlen` bzw. `Fallzahlen/Saarland` abgelegt. Für jeden Standard, der Fallzahlen enthalten soll, müssen auch die entsprechenden Bevölkerungszahlen verfügbar sein. Außerdem wird für jeden Standard mit Fallzahlen für die Fallzahlen ein gesondertes Verzeichnis angelegt. (In diesem Fall existieren nur Fallzahlen für den Standard Saarland.) Die Fallzahlen werden mit dem Skript `Fallzahlen.loader` aus der Datei `neuerkrankungen_saar.data` geladen.

B.3.1.4 Laden der Kodierungstabellen

Als letztes werden die Kodierungstabellen geladen. Das Skript `loader` im Verzeichnis `Kodierungen` lädt alle Kodierungstabellen. In der Steuerdatei `kodierung.ctl` sind sowohl die Daten als auch die Formatinformationen für alle Kodierungstabellen bis auf die im weiteren Verlauf gesondert behandelten Tabellen gespeichert. Viele dieser Tabellen beinhalten nur wenige Tupel, so daß es ein unnötiger Aufwand wäre, für jede dieser kleinen Relationen eigene Steuer- und Datendateien anzulegen.

Die Diagnose-Relationen sowie die Berufe und Staaten sind nicht in dieser Steuerdatei integriert, sondern werden durch eigene Loader-Skripte geladen.

Histologien

Die Histologien werden durch das Skript `Hist.loader` aus dem Verzeichnis `Diagnosen/Histologien` geladen. Erst werden die zu speichernden Histologie-Versionen in die Datenbank eingetragen, und danach werden jeweils durch eigene Steuerdateien die Ausprägungen der Histologien in die Datenbank eingefügt.

ICD-Kodes

Die ICD-Kodes werden durch das Skript `ICD.loader` aus dem Verzeichnis `Diagnosen/ICD_Codes` geladen. Dieses Skript trägt zuerst die verfügbaren ICD-Versionen ein. Da alle ICD-Kodes einer ICD-Version in einer Textdatei stehen, müssen die Codes erst in 3- und 4-stellige Codes aufgeteilt werden, damit diese getrennt geladen werden können.

Zusätzlich gibt es noch Dateien mit eigenen Überarbeitungen der beschreibenden Texte. Diese Texte ersetzen, sofern sie verfügbar sind, die Texte der originalen ICD-Kodes. Die eigenen ICD-Kodes (für alle Krebserkrankungen) werden zunächst in eine temporäre Tabelle geladen und erst danach in die Tabelle mit allen ICD-Kodes eingefügt.

Außerdem werden die 3-stelligen ICD-Kodes als 4-stellig übernommen, die ausschließlich als 3-stellige ICD-Kodes verfügbar sind. Wenn beispielsweise kein ICD-Kode mit der Nummer 140.X (X in 0,...,9) existiert, so müßte der ICD-Kode 140 als 4-stelliger Kode aufgenommen werden. Wenn aber beispielsweise der Kode 140.1 als 4-stelliger existieren würde, so wäre dies nicht notwendig.

Lokalisationen

Die Lokalisationen werden durch das Skript `Lok.loader` aus dem Verzeichnis `Diagnosen/Lokalisationen` geladen. Erst werden die zu speichernden Lokalisations-Versionen in

die Datenbank eingetragen, und danach werden jeweils durch eigene Steuerdateien die Ausprägungen der Lokalisationen in die Datenbank eingefügt.

Versionen

Die aktuell in der Relation `Erweiterte_Tumordaten` benutzten Diagnoseversionen werden durch das Skript `Versionen.loader` aus dem Verzeichnis `Diagnosen/Versionen` in die Relation `K_Versionen` geladen. Da Fremdschlüsselbedingungen auf die ICD-Kodes und Histologien existieren, muß diese Tabelle immer nach dem Ausführen der anderen Loader-Skripte geladen werden. Ebenso muß die Tabelle gelöscht werden, bevor die ICD-Kodes oder die Histologien neu geladen werden können.

Berufe und Staaten

Nach dem Laden der Diagnosen werden noch die Berufe bzw. Staaten mit der Skripten `Berufe.loader` bzw. `Staaten.loader` aus den Verzeichnissen `Berufe` bzw. `Staaten` wie üblich geladen.

B.3.2 Änderungen an den geladenen Daten

Prinzipiell sollen in der Datenbank keine Änderungen per SQL-Befehl direkt vorgenommen werden, da es Abhängigkeiten zwischen den Daten gibt, die unbedingt beachtet werden müssen. Damit der Benutzer trotzdem Daten in seiner Datenbank ändern kann, werden einige Skripte in dem Verzeichnis `data` bereitgestellt, die Änderungen an den gebräuchlichsten Daten sicher und einfach vornehmen. Mit diesen Skripten können die folgenden Daten nachträglich ersetzt bzw. modifiziert werden:

- Einwohnerdaten
- ICD-Kodes
- Histologien
- Lokalisationen
- Benutzte ICD- und Histologie-Version

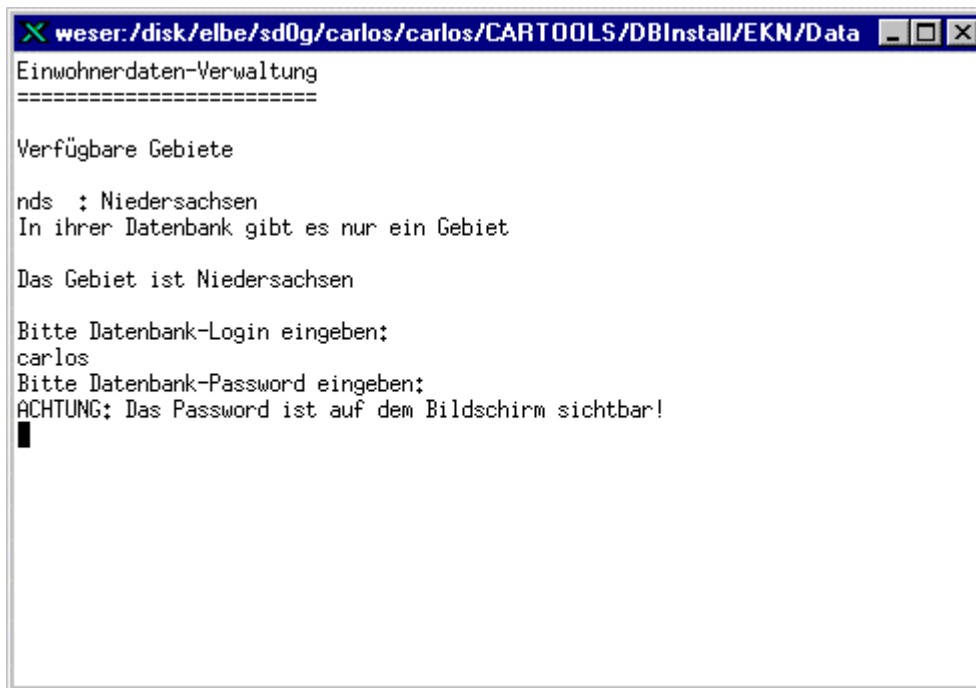
Das Laden der Falldaten ist abhängig von der Quelle der Daten. Wenn das vollständige Registerstellenschema existiert, dann werden die Falldaten aus den entsprechenden Relationen extrahiert (siehe Abschnitt B.4.1). Es ist aber auch möglich, daß nur das Auswertungssystem CARESS ohne eine vollständige Registerstelleninstallation benutzt werden soll. Für diesen Fall müssen die Falldaten aus einer anderen Quelle geladen werden. Insgesamt resultiert daraus, daß kein einheitliches Skript zum Laden der Falldaten angeboten werden kann.

Derzeit können neue ICD-, Histologie- und Lokalisations-Versionen nicht dynamisch hinzugefügt werden. Die Namen und Anzahl der Dateien mit den entsprechenden Kodierungen stehen fest, in zukünftigen Versionen der Installationsskripte werden diese Beschränkungen aufgehoben werden.

B.3.2.1 Einwohnerdaten

Das Skript zum Laden der Einwohnerdaten besitzt eine einfache eigene Benutzerschnittstelle, da hier eine Vielzahl von Möglichkeiten besteht, die sich über Kommandozeilenparameter nicht mehr einfach realisieren lassen.

Nach dem Start des Skripts `Einwohner-Loader` erscheint die folgende Maske auf dem Bildschirm:



```
weser:/disk/elbe/sd0g/carlos/carlos/CARTTOOLS/DBInstall/EKN/Data
Einwohnerdaten-Verwaltung
=====

Verfügbare Gebiete

nds : Niedersachsen
In ihrer Datenbank gibt es nur ein Gebiet

Das Gebiet ist Niedersachsen

Bitte Datenbank-Login eingeben;
carlos
Bitte Datenbank-Passwort eingeben;
ACHTUNG: Das Passwort ist auf dem Bildschirm sichtbar!
█
```

Zunächst muß ausgewählt werden, für welches Gebiet Einwohnerdaten geladen werden sollen. Zu diesem Zweck muß das jeweils vor dem Doppelpunkt stehende Kürzel (z.B. `nds` für Niedersachsen) eingegeben werden. Wenn nur ein Gebiet in der Datenbank ist, dann wählt das Skript automatisch dieses eine Gebiet aus.

Nach der Wahl des Gebiets müssen der Datenbank-Benutzername und das zugehörige Passwort eingegeben werden. Hierbei ist zu beachten, daß das Passwort als Klartext auf dem Bildschirm erscheint.

Nach der Eingabe dieser Parameter können mit dem Programm die Einwohnerdaten modifiziert werden. Die Auswahl, welche Operation ausgeführt werden soll, geschieht durch die Angabe des entsprechenden vor dem Doppelpunkt stehenden Kürzels der Operation:

- Mit der Operation `add` können einzelne Jahrgänge aus den Dateien mit den Einwohnerdaten in die Datenbank geladen werden. Es werden nur die Jahrgänge der Einwohnerdaten zur Auswahl angeboten, die noch nicht in die Datenbank geladen wurden. Es muß jeweils eine Datei mit den Einwohnermetainfos und eine dazugehörige Datei mit den Einwohnerzahlen übergeben werden.
- Mit der Operation `all` werden alle Jahrgänge aus einer Datei mit Einwohnerdaten in die Datenbank geladen. Jahrgänge die bereits in der Datenbank sind, werden nicht geladen und somit nicht modifiziert.
- Mit der Operation `copy` kann ein Jahrgang aus der Datenbank für einen anderen Jahrgang kopiert werden.

- Mit der Operation `del` wird ein Jahrgang aus der Datenbank gelöscht.
- Mit der Operation `list` werden alle in der Datenbank befindlichen Jahrgänge aufgelistet.
- Die Operation `quit` beendet das Programm.

```
weser:/disk/elbe/sd0g/carlos/carlos/CARTTOOLS/DBInstall/EKN/Data
Einwohnerdaten Verwaltung für Niedersachsen
=====
add : Neue Jahrgänge von Einwohnerdateien in die Datenbank laden
all : Neue Einwohnerdateien in die Datenbank laden
copy : Einwohnerdateien in der Datenbank kopieren
del : Einwohnerdaten aus der Datenbank löschen
list : Alle Jahrgänge anzeigen
quit : Programm beenden

Kommando?
█
```

Jede Änderung in den Einwohnerdaten wird so vermerkt, daß alle Änderungen bei einem eventuell nachfolgenden vollständigen Laden der Daten (siehe Abschnitt B.3.1) rekonstruiert werden. Die Einwohnerzahlen der einzelnen Jahrgänge in der Datenbank z.B. für das Gebiet Niedersachsen werden in das Unterverzeichnis `Einwohnerdaten.Niedersachsen` kopiert.

Zum Einlesen von Einwohnerzahlen in die Datenbank werden zwei Dateien benötigt. Die eine Datei enthält Daten über die Struktur der Einwohnerzahlen (Einwohnermetainfos) und die andere Datei enthält die Daten selbst. Die Dateien mit den Einwohnerzahlen können mehrere Jahre enthalten. Ein Jahrgang muß jedoch in einer Datei immer vollständig enthalten sein. Die Datei mit den Einwohnermetainfos enthält für ein Untersuchungsgebiet die Gruppeneinteilungen der Einwohnerzahlen nach Altergruppe, Geschlecht, Jahr. Die Datei muß dazu die Werte der Attribute *untobjnr*, *jahr*, *geschlecht*, *altermin* und *altermax* jeweils in einer Zeile durch das Zeichen `|` getrennt enthalten. Die Datei mit den Einwohnerzahlen enthält eine Einwohnerzahl für jedes Gebiet und jede Kombination von Alter, Geschlecht und Jahr im jeweiligen Untersuchungsgebiet, die in den Einwohnermetainfos festgelegt ist. Diese Zahlen werden in die Relation Einwohner geladen. Die Datei mit den Einwohnerzahlen muß dazu die Werte der Attribute *altermin*, *geschlecht*, *jahr*, *untobjnr*, *gebiet* und *einwohnerzahl* jeweils in einer Zeile durch das Zeichen `„|„` getrennt enthalten.

Die Formate der Eingabedateien können auch in den Readme-Dateien für die Einwohnerzahlen und Einwohnermetainfos in dem Verzeichnis `README` nachgelesen werden.

B.3.2.2 ICD-Kodes laden

Die ICD-Kodes können durch das Skript `ICD-Loader` nachträglich geändert werden. Dieses Skript verlangt den Benutzernamen und das Passwort als Parameter:

```
ICD-Loader <DB-Benutzer> <DB-Passwort>
```

Die Eingabedateien für dieses Skript sind:

<code>icd_9_3-krebs.txt</code>	3-stellige ICD-Kodes nach ICD 9
<code>icd_9_4-krebs.txt</code>	4-stellige ICD-Kodes nach ICD 9
<code>icd_10_3-krebs.txt</code>	3-stellige ICD-Kodes nach ICD 10
<code>icd_10_4-krebs.txt</code>	4-stellige ICD-Kodes nach ICD 10
<code>ICDVersion.sql</code>	alle verfügbaren ICD-Versionen

Jede Zeile in den ICD-Dateien stellt einen 3- bzw. 4-stelligen ICD-Kode mit einer textuellen Beschreibung dar. Das Format aller in dieser Datei befindlichen Zeilen ist daher identisch und lautet

- `<3 Zeichen ICD-Kode><Semikolon><Maximal 120 Zeichen Beschreibung>`
für einen 3-stelligen ICD-Kode bzw.
- `<3 Zeichen ICD-Kode><Punkt><4. Stelle><Semikolon><Maximal 120 Zeichen Beschreibung>`
für einen 4-stelligen ICD-Kode.

Die Datei `ICDVersion.sql` weicht von diesem Format ab und enthält pro ICD-Version ein SQL-INSERT-Kommando für die Relation `K_ICD_VERSION` mit dem Namen und der Kennung der Version.

Das Format dieser Dateien ist in den zugehörigen `readme`-Dateien im Unterverzeichnis `README` nachzulesen.

B.3.2.3 Histologien laden

Die Histologien können durch das Skript `Histologie-Loader` nachträglich geändert werden. Dieses Skript verlangt den Benutzernamen und das Passwort als Parameter:

```
Histologie-Loader <DB-Benutzer> <DB-Passwort>
```

Die Eingabedateien für dieses Skript sind:

<code>hist_o-da.txt</code>	ICD-O-DA 1
<code>hist_vol.txt</code>	ICD-O 1
<code>hist_vo2.txt</code>	ICD-O 2
<code>HistologieVersion.sql</code>	alle verfügbaren Histologie-Versionen

Jede Zeile in den Dateien `hist_o-da.txt`, `hist_vol.txt` bzw. `hist_vo2.txt` stellt einen Histologie-Kode mit einer textuellen Beschreibung dar. Das Format aller in diesen Datei

befindlichen Zeilen ist daher identisch, steht in der Datei `hist_data.readme` im Unterverzeichnis `README` und lautet wie folgt:

```
<4 Zeichen Histologie-Kode><Schrägstrich><Dignität><Semikolon><Maximal 120 Zeichen Beschreibung>
```

Die einzige Ausnahme bildet wie bei den ICD-Kodes die Datei `HistologieVersion.sql`, in der pro Histologie-Version ein SQL-INSERT-Kommando steht, das deren Namen und Nummer einträgt.

B.3.2.4 Lokalisationen laden

Die Lokalisationen können durch das Skript `Lokalisation-Loader` nachträglich geändert werden. Dieses Skript verlangt den Benutzernamen und das Passwort als Parameter:

```
Histologie-Loader <DB-Benutzer> <DB-Passwort>
```

Die Eingabedateien für dieses Skript sind:

<code>lok_v3.txt</code>	3. Auflage (ICD-O-1)
<code>lok_v5.txt</code>	5. Auflage (ICD-O-2)
<code>LokalisationVersion.sql</code>	alle verfügbaren Lokalisations-Versionen

Jede Zeile in den Lokalisations-Dateien stellt einen 3- bzw. 4-stelligen Lokalisations-Kode mit einer textuellen Beschreibung dar. Das Format aller in dieser Datei befindlichen Zeilen ist daher identisch und lautet

- `<3 Zeichen Lokalisations-Kode><Semikolon><Maximal 120 Zeichen Beschreibung>`
für einen 3-stelligen Lokalisations-Kode bzw.
- `<3 Zeichen Lokalisations-Kode><Punkt><4. Stelle><Semikolon><Maximal 120 Zeichen Beschreibung>`
für einen 4-stelligen Lokalisations-Kode.

Die Datei `LokalisationVersion.sql` weicht von diesem Format ab und enthält pro Lokalisations-Version ein SQL-INSERT-Kommando für die Relation `K_Lokalisation_Version` mit dem Namen und der Kennung der Version.

Das Format dieser Dateien ist in den zugehörigen `readme`-Dateien im Unterverzeichnis `README` nachzulesen.

B.3.2.4 Aktuelle Diagnoseversionen festlegen

In der Datei `DiagnoseVersionen.sql` steht ein SQL-INSERT-Kommando, daß die aktuelle Versionsnummern der benutzten ICD- und Histologie-Versionen in der Datenbank in die Relation `K_Versionen` einträgt. Die jeweils angegebenen Nummern müssen auch in den Tabellen `K_ICD_Version` bzw. `K_Histologie_Version` verfügbar sein. Alle Falldaten in der Relation `Erweiterte_Tumordaten` müssen mit diesen Kodierungsversionen kodiert sein. Wenn Tupel in der Relation `Erweiterte_Tumordaten` mit anderen Versionen kodiert sind, liefert das Auswertungsprogramm CARESS fehlerhafte Ergebnisse. Dieses Skript wird immer dann automatisch aufgerufen, wenn die ICD- oder die Histologie-Dateien neu eingelesen werden.

B.4 Sonstige Verzeichnisse

In diesem Abschnitt finden sich Informationen zu allen sonstigen Verzeichnissen, die nicht direkt mit der Installation des Schemas oder der Daten zu tun haben. Dazu gehört das Verzeichnis mit dem Skript zum Füllen der Erweiterten_Tumordaten aus der Registerstellendatenbank (Abschnitt B.4.1) und in Abschnitt B.4.2 ein Verzeichnis mit einigen Hilfsskripten, die die Arbeit vereinfachen sollen.

B.4.1 Füllen der Relation Erweiterte_Tumordaten

Im Verzeichnis `Erweiterte_TD_fuellen` befindet sich das SQL-Skript `fill_erweiterte_tumordaten.sql`, das Meldungen in die Relation `Erweiterte_Tumordaten` überträgt. Es werden nur die Patienten bzw. die Tumoren dieser Patienten übertragen, die seit dem letzten Transfer geändert worden sind. Das Auswertungsprogramm CARESS bezieht die Falldaten ausschließlich aus der Relation `Erweiterte_Tumordaten`, aus diesem Grund beeinflusst die Arbeitsweise des Skripts `fill_erweiterte_tumordaten.sql` jede Auswertung. Die Funktionsweise des Skripts wird in diesem Abschnitt erklärt werden.

CARELIS trägt für jeden Patienten, der neu ist oder verändert wurde, dessen Patient-Id (Primärschlüssel von Patienten) in die Relation `Geaenderte_Patienten` ein. Das Skript löscht nun alle Tumoren der Patienten, die verändert wurden, mit Hilfe der Patient-Id aus der Relation `Erweiterte_Tumordaten`. Danach werden alle Tumoren der geänderten Patienten neu in diese Relation eingefügt. Wenn man die Ids aller existierenden Patienten in die Relation `Geaenderte_Patienten` einträgt, so wird die Relation `Erweiterte_Tumordaten` vollständig neu erzeugt.

Die zum Erzeugen neuer Tupel benötigten Daten liegen in den Relationen `Tumoren`, `Patienten` und `Sterbefaelle`. Diese Informationen werden über einen Join aus diesen Relationen zusammengeführt:

```
SELECT      ...
FROM        Geaenderte_Patienten d, Tumoren a,
            Patienten b, Sterbefaelle c
WHERE       d.PATIENT_ID = b.PATIENT_ID AND
            a.PATIENT_REF = b.PATIENT_ID AND
            a.PATIENT_REF = c.PATIENT_ID (+);
```

Da nicht jeder Patient ein Sterbefall ist, ist der Join bezüglich der Relation `Sterbefaelle` ein „outer Join“, d.h. falls kein Sterbefall zu einem Patienten existiert, so ist die Join-Bedingung trotzdem erfüllt und alle Informationen, die aus dem Sterbefall resultieren würden, werden mit NULL-Werten erzeugt.

Es werden nur Tumoren übernommen, die nach den aktuellen Kodierungsversionen kodiert sind, die in der Relation `K_Versionen` vermerkt sind. (Wenn ein Tumor beispielsweise nach einer anderen ICD-Version als in der Relation `K_Versionen` kodiert wurde, so wird dieser Tumor nicht in die Relation `Erweiterte_Tumordaten` übernommen.) Das Ziel dieser Prüfung ist es, einheitliche Kodierungsversionen in der Relation `Erweiterte_Tumordaten` zu erhalten, um Auswertungen mit CARESS über diesen Datenbestand durchführen zu können. Das hat zur Folge, daß wenn die Kodierungsversionen in der Relation `K_Versionen` geändert werden, die Relation `Erweiter-`

te_Tumordaten vollständig neu erzeugt werden muß. Dieser Vorgang ist zur Zeit nicht durch ein Skript abgedeckt, da ein Wechsel der Kodierungsversion im allgemeinen recht selten ist.

In den Relationen Tumoren, Patienten und Sterbefälle stehen für jedes Ereignis nur die Zeitpunkte, d.h. das Jahr und der Monat. Diese Angaben können unter Umständen auch unbekannt sein. In der Relation Erweiterte_Tumordaten werden aber auch das Diagnose- bzw. Sterbealter gespeichert. Aus diesem Grund werden aus den obigen Zeitpunkten die benötigten Altersangaben berechnet. Die Berechnung ist für das Diagnose- bzw. Sterbealter analog und wird hier nur exemplarisch für das Diagnosealter erklärt. Wenn sowohl das Jahr der Geburt als auch das Jahr der Diagnose unbekannt ist, dann ist auch das Diagnosealter unbekannt. Wenn einer der beiden Monate unbekannt sind, dann ergibt sich das Diagnosealter aus der Differenz „*diagnose_jahr – geburtsjahr*“. Nur wenn sowohl der Geburtsmonat als auch der Diagnosemonat bekannt ist, wird das Diagnosealter entweder, wenn der Diagnosemonat kleiner als der Geburtsmonat ist, durch „*diagnose_jahr – geburtsjahr – 1*“ oder ansonsten durch „*diagnose_jahr – geburtsjahr*“ bestimmt. Der *Diagnosezeitpunkt* wird aus dem *Diagnosedatum* berechnet (siehe Abschnitt 2.2.7).

Weiterhin werden in dem Attribut *wohnort* Gemeindecodenziffern benutzt. In der Relation Erweiterte_Tumordaten werden aber spezielle interne Codes zur Kodierung der Regionen benötigt. Die Gemeindecodenziffern werden mit Hilfe der Relation Gebiete in diese internen Codes umgesetzt.

B.4.2 Hilfsskripte

In dem Verzeichnis *Skripte* befinden sich einige Hilfsskripte, mit deren Hilfe Verwaltungsaufgaben wahrgenommen werden können. Das Skript *FilterConstraints* in diesem Verzeichnis dient dazu, alle Constraints mit einem bestimmten Text im Bedingungsteil zu suchen und zu löschen. Dies wird dann notwendig, wenn sich beispielsweise der Check-Constraint eines Attributs geändert hat. Mit diesem Skript können alle Constraints mit der alten Bedingung gelöscht werden, und danach kann man die neuen Constraints einfügen.

Literaturverzeichnis

- [AFH+97] H.-J. Appelrath, J. Friebe, E. Hinrichs, H. Hinrichs, I. Hoting, J. Kieschke, J. Rettig, A. Scharnofske, W. Thoben, and F. Wietek. CARLOS (Cancer Registry Lower-Saxony): Tätigkeitsbericht für den Zeitraum 1.1.-31.12.1997. Technischer Bericht, OFFIS, Oldenburg, Dezember 1997.
- [AFH+96] H.-J. Appelrath, J. Friebe, H. Hinrichs, V. Kamp, J. Rettig, W. Thoben, and F. Wietek. CARLOS (Cancer Registry Lower-Saxony): Tätigkeitsbericht für den Zeitraum 1.1.-31.12.1996. Technischer Bericht, OFFIS, Oldenburg, Dezember 1996.
- [AFK+95] H.-J. Appelrath, J. Friebe, V. Kamp, J. Rettig, W. Thoben, and F. Wietek. CARLOS (Cancer Registry Lower-Saxony): Tätigkeitsbericht für den Zeitraum 1.1.-31.12.1995. Technischer Bericht, OFFIS, Oldenburg, Dezember 1995.
- [AdV91] Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV). *ATKIS-Gesamtdokumentation, Teil D, ATKIS-Objektartenkatalog*. Niedersächsisches Landesvermessungsamt, Hannover, 1991.
- [KRG94] Deutscher Bundestag. Gesetz über Krebsregister (Krebsregistergesetz KRG). *Bundesgesetzblatt*, I(79): 3351-3355, November 1994.
- [DIM93] DIMDI, editor. *ICD-9: Internationale Klassifikation der Krankheiten, Verletzungen und Todesursachen (ICD) in der Fassung der vom Bundesministerium für Gesundheit herausgegebenen 9. Revision. Zusätzliche Systematiken und Klassifizierungsregeln*, volume I Teil B of . Deutsches Institut für Medizinische Dokumentation und Information (DIMDI). Verlag W. Kohlhammer, Köln, 1993.
- [DIM95] DIMDI, editor. *ICD-10: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10. Revision*, volume 1 - Systematisches Verzeichnis of . Deutsches Institut für Medizinische Dokumentation und Information (DIMDI) im Auftrag des Bundesministeriums für Gesundheit. ecomed Verlagsgesellschaft, Landsberg, 1995.
- [EWLP95] D. Esteban, S. Whelan, A. Laudico, and D. M. Parkin. Manual for Cancer Registry Personnel. IARC Technical Report 10, International Agency for Research on Cancer (IARC), World Health Organization (WHO) and International Association of Cancer Registries (IACR), Lyon, 1995.
- [Fri97] J. Friebe, editor. *Eine GeoServer-Architektur zur Bereitstellung geographischer Basisdaten im Internet, BTW '97*, Ulm, 1997.
- [Her78] Hermanek, editor. *Tumor-Histologie-Schlüssel - ICD-O-DA, 2. Auflage*. Springer Verlag, Berlin, 1 edition, 1978.
- [HSSW92] P. Hermanek, O. Scheibe, B. Spiessl, and G. Wagner, editors. *TNM-Klassifikation maligner Tumoren, 2. Revision*. Springer Verlag, Berlin, Heidelberg, 4 edition, 1992.
- [HDW97] H. Hinrichs, J.-C. Dirks, and T. Wilkens. Record Linkage im Epidemiologischen Krebsregister Niedersachsen. , OFFIS, 1.12.1997 1997. In Vorbereitung.
- [KW75] S. Koller and G. Wagner. *Handbuch der medizinischen Dokumentation und Datenverarbeitung*. F. K. Schattauer Verlagsgesellschaft, Stuttgart, 1975.

- [KSSM94] A. Krtschil, I. Schmidtman, J. Schüz, and J. Michaelis. Bericht über die Pilotstudie zum Krebsregister Rheinland-Pfalz. Technischer Bericht, Tumorzentrum Rheinland-Pfalz, Mainz, Dezember 1994.
- [LGH95] F. Leiner, W. Gaus, and R. Haux. *Medizinische Dokumentation: Einführendes Lehrbuch*. F. K. Schattauer Verlagsgesellschaft, Stuttgart, 1995.
- [PCF+94] D. M. Parkin, V. W. Chen, J. Ferlay, J. Galceran, H. H. Storm, and S. L. Whelan. Comparability and Quality Control in Cancer Registration. IARC Technical Report 19, International Agency for Research on Cancer (IARC), Lyon, 1994.
- [PVM90] C. Percy, V. Van Holten, and C. Muir, editors. *ICD-O - International Classification of Diseases for Oncology*. World Health Organization, Genf, 2 edition, 1990.
- [Sch96] J. Schüz. Dokumentationsstandards in epidemiologischen Krebsregistern nach dem Krebsregistergesetz (KRG). Arbeitspapier, Arbeitsgemeinschaft bevölkerungsbezogener Krebsregister in Deutschland, Mainz, Februar 1996.
- [SAW95] J. Schüz, U. Altmann, and W. Wächter. Vorschlag für Dokumentationsstandards in epidemiologischen Krebsregistern nach dem Krebsregistergesetz (KRG). Arbeitspapier, GMDS-Arbeitsgruppe "Dokumentationsstandards für epidemiologische Krebsregister", Mainz, Gießen, März 1995.
- [SBu92] Statistisches Bundesamt. Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbenennungen, Wiesbaden, 1992.
- [Wag88] G. Wagner, editor. *Tumorlokalisierungsschlüssel. International Classification of Diseases for Oncology ICD-O. Topographischer Teil*. Springer Verlag, Berlin, Heidelberg, 3 edition, 1988.
- [Wag93] G. Wagner, editor. *Tumorlokalisierungsschlüssel. International Classification of Diseases for Oncology ICD-O, 2. Auflage. Topographischer Teil*. Springer Verlag, Berlin, 5 edition, 1993.
- [Wie96] F. Wietek. Das epidemiologische Datenschema des Niedersächsischen Krebsregisters für die Erprobungsphase 1995-97. Technischer Bericht, OFFIS, Oldenburg, April 1996.
- [WS97] F. Wietek and A. Scharnofske. Das epidemiologische Datenschema des Niedersächsischen Krebsregisters für die Erprobungsphase 1995-97. Technischer Bericht, OFFIS, August 1997.
- [WW97] C. Wittekind and G. Wagner. *TNM Klassifikation maligner Tumoren, 5. Auflage*, volume 5 of . Springer, 1997.
- [WHO76] World Health Organization, editor. *ICD-O - International Classification of Diseases for Oncology*. WHO, Genf, 1 edition, 1976.
- [ZUMA85] ZUMA. Vercodung offener Berufsangaben. Praktische Hinweise. Bericht 85/11, Zentrum für Umfragen, Methoden und Analysen e.V. (ZUMA), Mannheim, September 1985.