Registerstelle des Krebsregisters Schleswig-Holstein

Institut für Krebsepidemiologie e.V.

Pseudonymisierungslösungen in den Krebsregistern Schleswig-Holstein und Nordrhein-Westfalen

Dipl.- Inf. Anke Richter

Dr. Volker Krieg

TMF Workshop ID-Management, Berlin, 15.12.2008

Ziele und Aufgaben der epidemiologischen Krebsregister

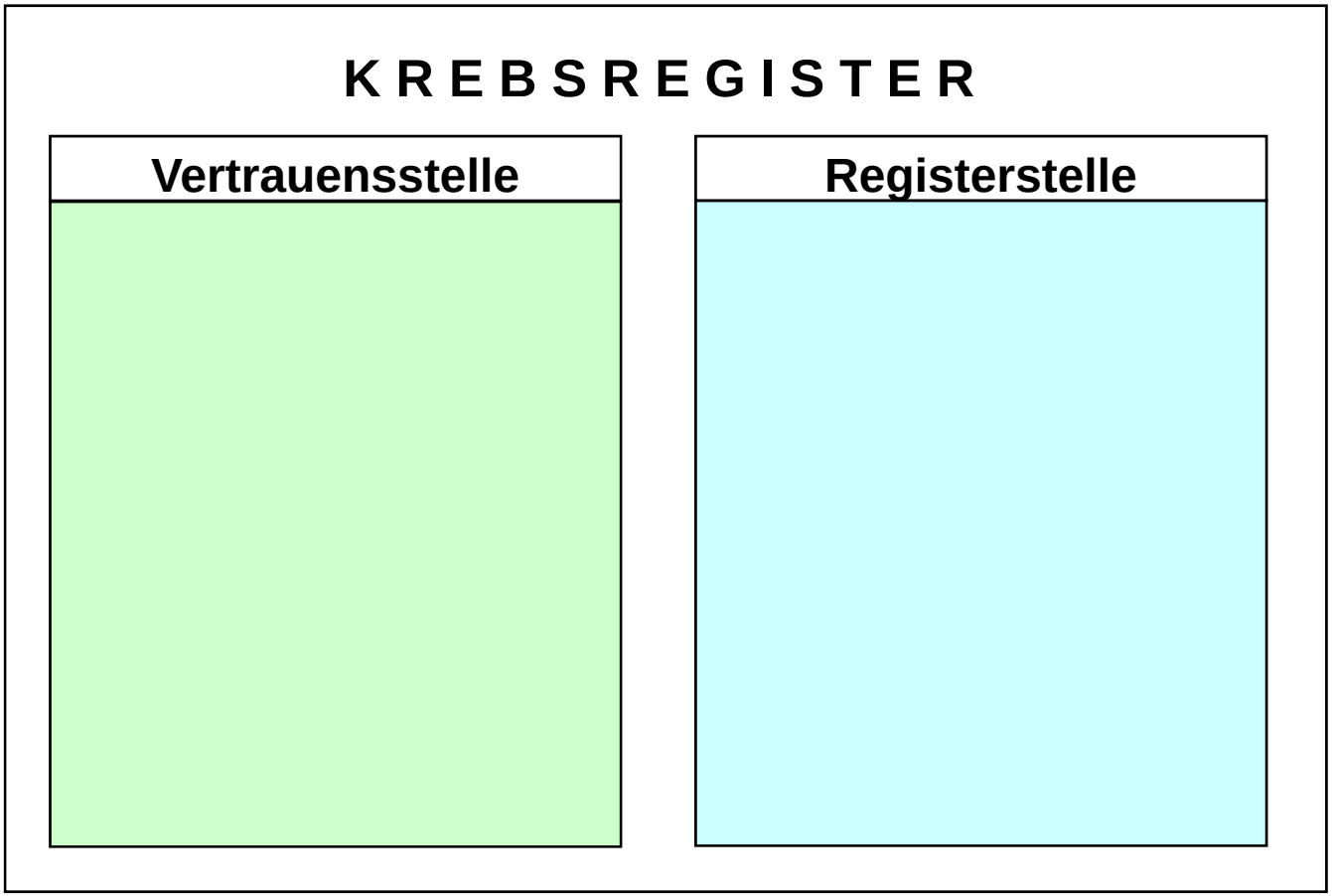
Krebsregister haben das Auftreten und die Trendentwicklung aller Formen von Krebserkrankungen zu beobachten, insbesondere statistisch-epidemiologisch auszuwerten, Grundlagen der Gesundheitsplanung sowie der epidemiologischen Forschung einschließlich der Ursachenforschung bereitzustellen und zu einer Bewertung präventiver und kurativer Maßnahmen beizutragen. Sie haben vornehmlich anonymisierte Daten für die wissenschaftliche Forschung zur Verfügung zu stellen.

§1 Bundeskrebsregistergesetz

Pseudonymisierungslösungen im Krebsregister

Basismodell der Krebsregistrierung:

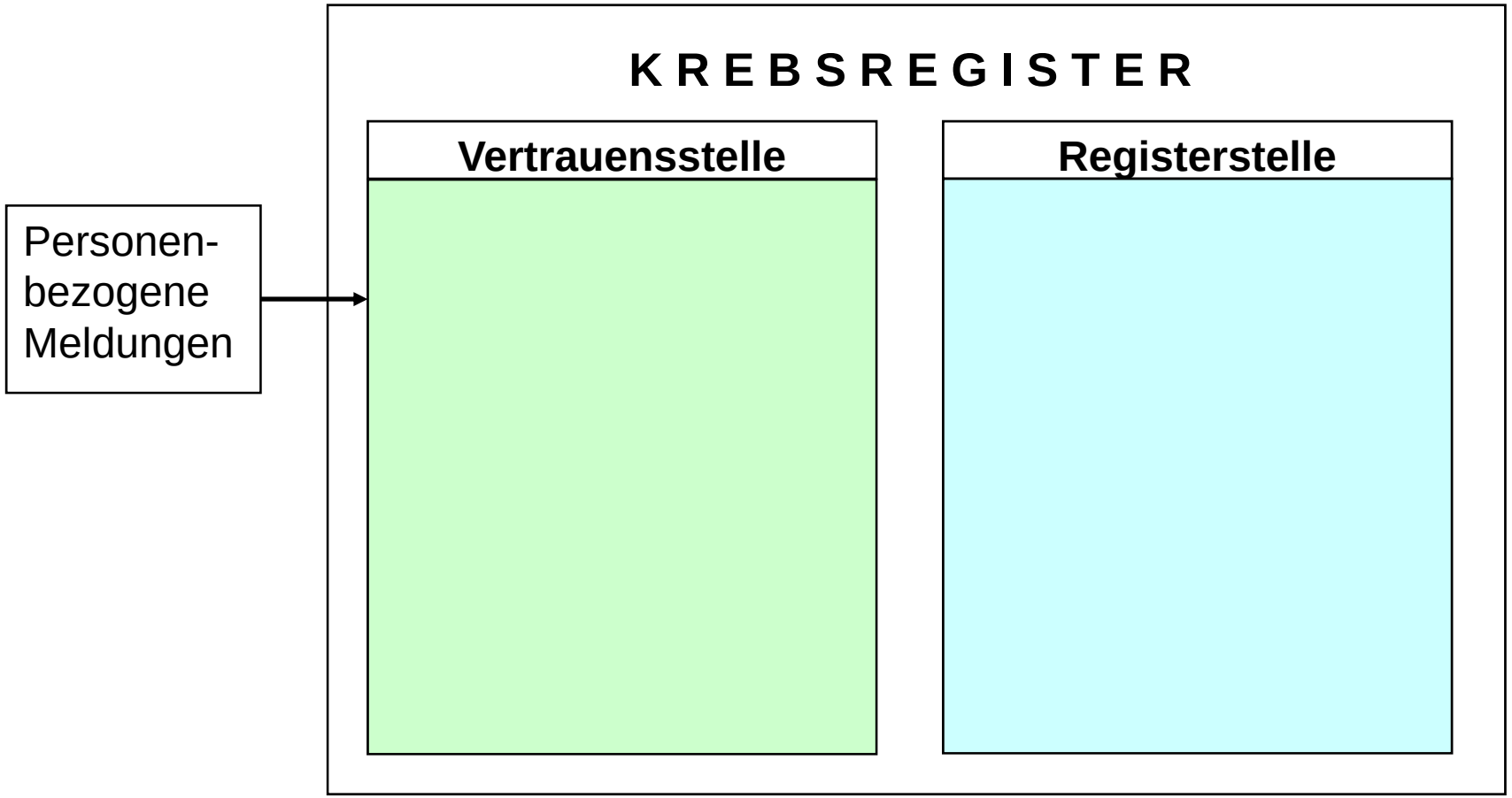
Zwei organisatorisch unabhängige Einheiten: Vertrauens- und Registerstelle



Pseudonymisierungslösungen im Krebsregister

Basismodell der Krebsregistrierung:

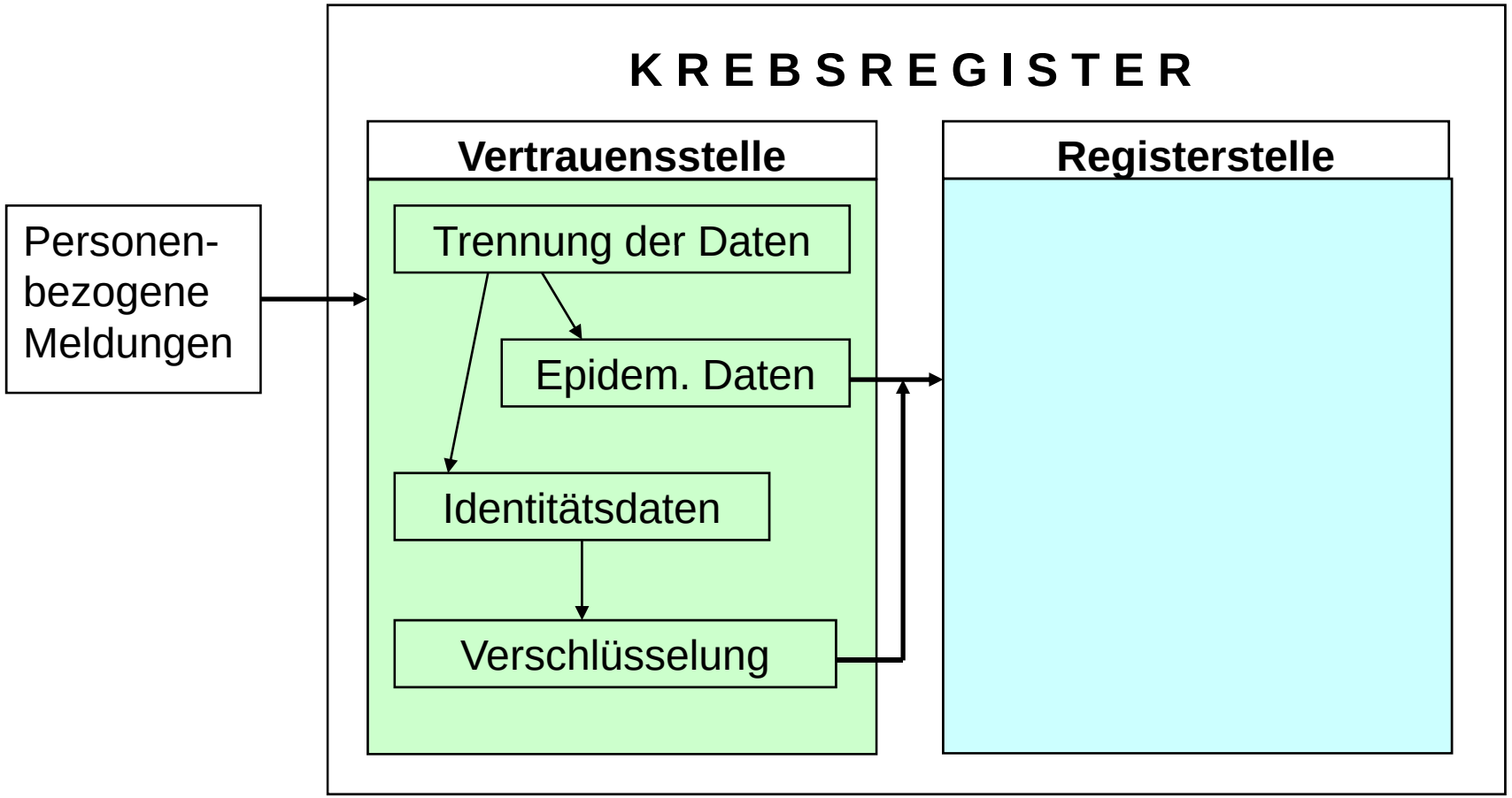
Meldungen aus unterschiedlichen Quellen werden in der Vertrauensstelle bearbeitet



Pseudonymisierungslösungen im Krebsregister

Basismodell der Krebsregistrierung:

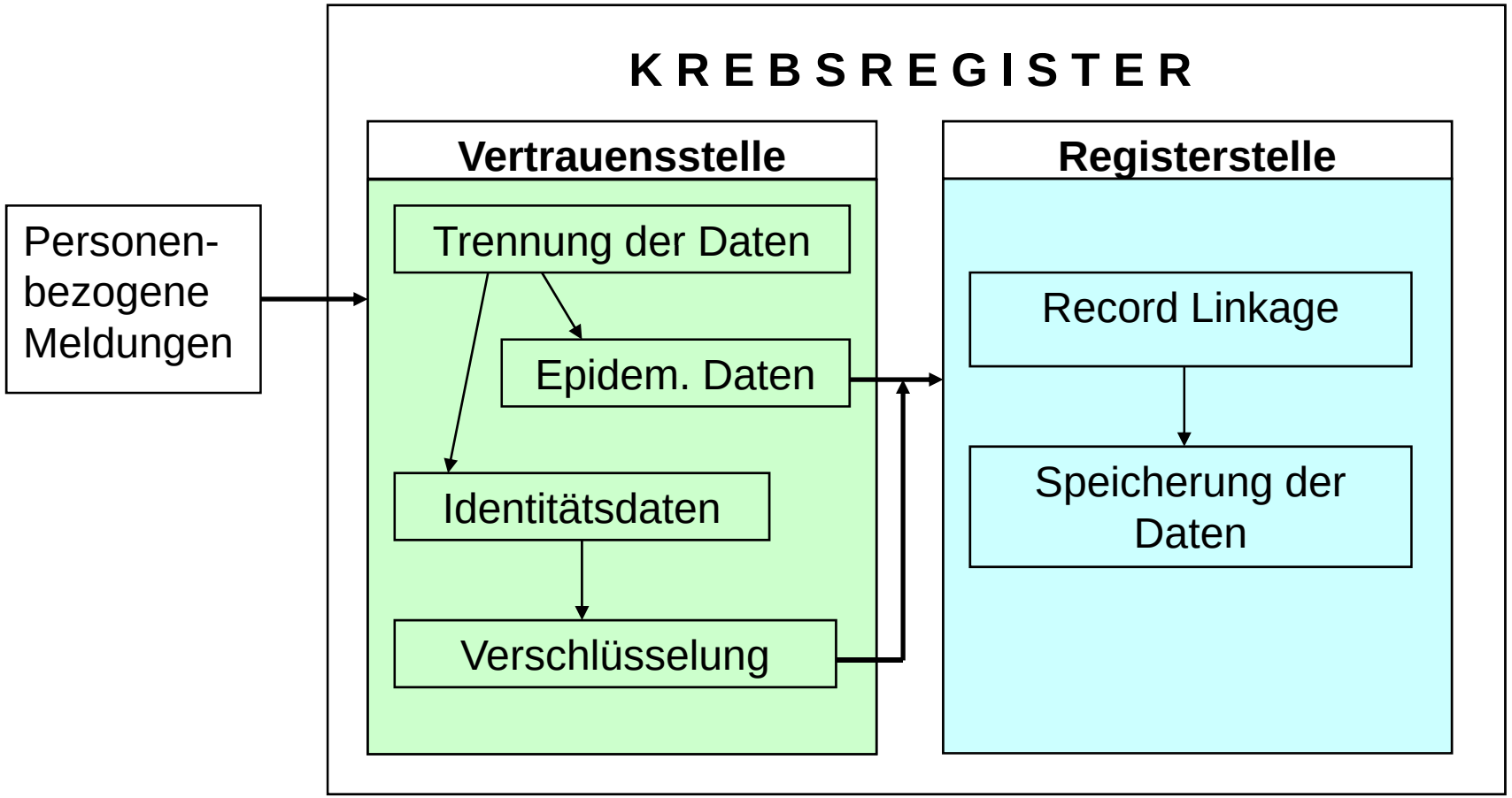
Bearbeitung, Trennung, Chiffrierung und Weiterleitung der Daten



Pseudonymisierungslösungen im Krebsregister

Basismodell der Krebsregistrierung:

Record Linkage und Speicherung in der Registerstelle



Chiffrierverfahren in Krebsregistern

Aufgabe:
Abgleich von Datensätzen aus unterschiedlichen Meldequellen

Auflage laut Gesetz:
Personenidentifizierende Daten dürfen nicht im Klartext gespeichert werden.

Lösung:
Es kommen zwei Verschlüsselungsverfahren unabhängig voneinander zur Anwendung:

- Asymmetrische Chiffrierung aller Identitätsdaten
- Bildung von sog. Kontrollnummern als Basis für den Abgleich

Asymmetrische Chiffrierung der Identitätsdaten mittels RSA-Verschlüsselung

Die personenidentifizierenden Daten werden in einer Zeichenkette zusammen gefasst und diese mit dem öffentlichen Schlüssel eines asymmetrischen RSA-Verschlüsselungsverfahrens chiffriert.

Weiterleitung an die Registerstelle, wo eine Dechiffrierung nicht möglich ist.

Der Schlüssel zum Dechiffrieren wird außerhalb des Krebsregisters aufbewahrt und nur unter besonderen gesetzlich geregelten Vorkehrungen extern zum Entschlüsseln verwendet, wenn es besondere Forschungsaufgaben erforderlich machen, um beispielsweise zusätzliche Daten zu einem bestehenden Kollektiv erheben zu können.

Kontrollnummern

Um neu eingehende Meldungen zu bereits registrierten Patienten zuzuordnen, ohne Klartextangaben von personenidentifizierenden Daten zu verwenden, sieht das Bundeskrebsregistergesetz sogenannte Kontrollnummern vor.

Kontrollnummer sind Verschlüsselungen von zuvor standardisierten Komponenten der personenidentifizierenden Daten einer Meldung (Name, Vorname, Geburtsname, früherer Name, Geburtstag, Titel).

Kontrollnummern eignen sich für einen Abgleich von Mehrfachmeldungen, sind jedoch als alleinige Abgleichskriterien nicht ausreichend. Nur wenn Klartextangaben wie Geschlecht, Geburtsmonat und –jahr und der Wohnort zusätzlich im Abgleich Berücksichtigung finden, werden zuverlässige Treffer- und Nichttrefferquoten erzielt.

Bildung von Kontrollnummern, Teil 1: Standardisierungsphase

- Umlaute und ß werden in zweibuchstabige Schreibweise umgesetzt
- Alles wird einheitlich groß geschrieben
- Name, Vorname, Geburtsname und früherer Name werden anhand von Trennzeichen in 3 Komponenten zerlegt
- Namenszusätze werden stets in der dritten Komponente abgelegt
- Zu jedem Namen wird seine phonetische Abbildung erzeugt (Kölner Phonetik), um Schreib- und Hörfehler zu kompensieren (Meier, Meyer, Maier, Mayer, Meuer, Mauer => MEER)

Pseudonymisierungslösungen im Krebsregister

Standardisierung:

Name:

Holz-Müller	→	HOLZ	MUELLER	
		HOLZMUELLER	ULSMELER	

Vorname:

Anne Marie	→	ANNE	MARIE	
		ANNEMARIE	EMEMERE	

Geburtsname:

Holz	→	HOLZ		
		HOLZ	ULS	

Früherer Name:

Meyer zur Alp	→	MEYER	ALP	ZUR
		MEYERALPZUR	MEERALPSUR	

Ergebnis der Zerlegung und Standardisierung aller Namensbestandteile

Komponenten	Ursprung
K1...K3	Nachname, 3-teilig (mit/ohne Namenszusätzen)
K4...K6	Vorname, 3-teilig
K7…K9	Geburtsname, 3-teilig
K10…K12	Früherer Name, 3-teilig
K13	Tag des Geburtsdatums
K14	DDR-Namenscode
K15	Phonetischer Code vom standardisierten Nachnamen
K16	Phonetischer Code vom standardisierten Vornamen
K17	Phonetischer Code vom standardisierten Geburtsnamen
K18	Phonetischer Code vom standardisierten früheren Namen
K19, K20	Titel, 2-teilig
K21, K22	Spezielle Komponenten für Baden-Württemberg

Bildung von Kontrollnummern, Teil 2: Chiffrierungsphase

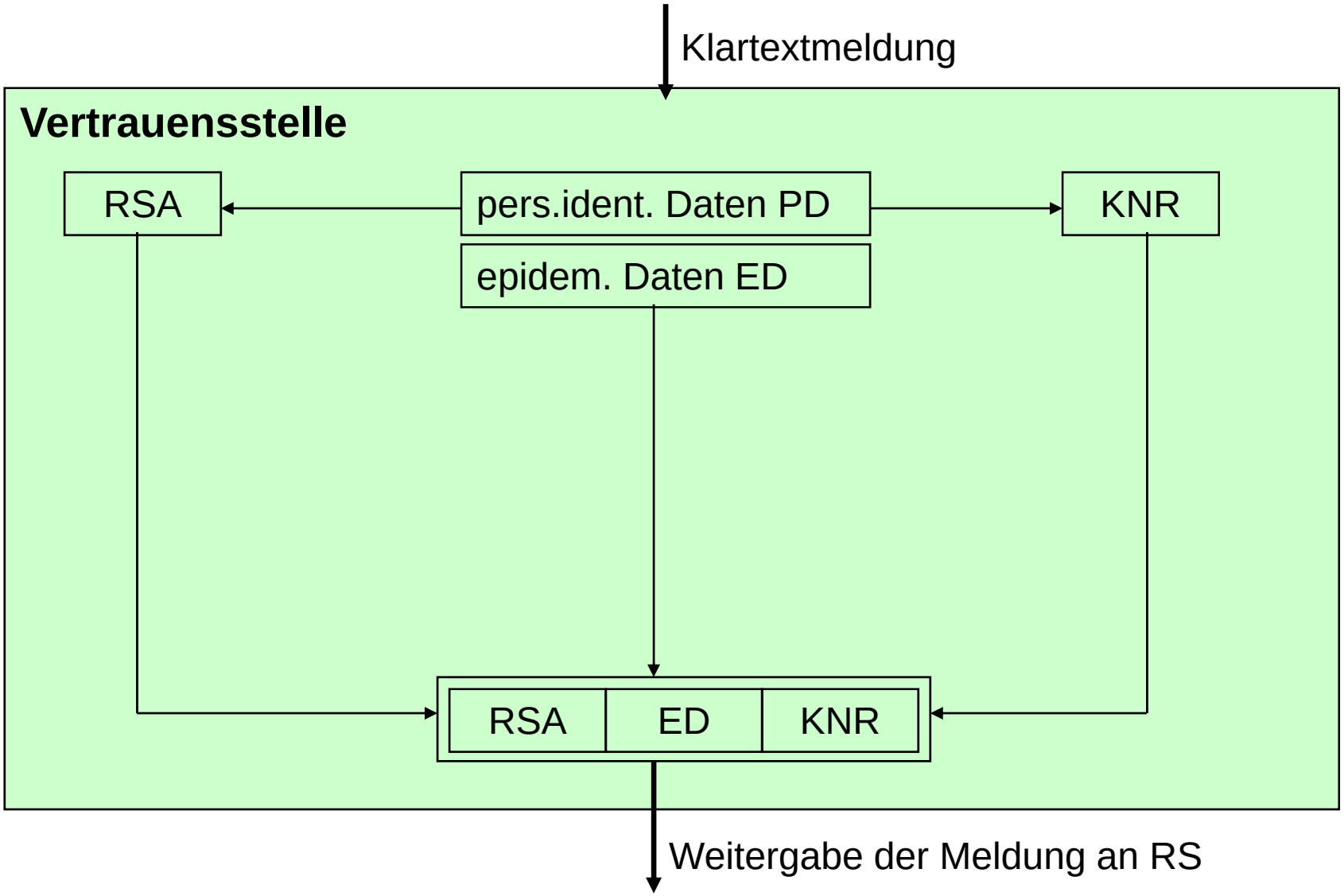
Wenn die Zerlegung der Namensbestandteile und deren Standardisierung abgeschlossen ist, schließt sich die Chiffrierungsphase an:

Auf jede der 22 Komponenten wird nacheinander ein
Einwegverschlüsselungsverfahren (MD5) und ein
symmetrisches Chiffrierverfahren (IDEA) angewendet.

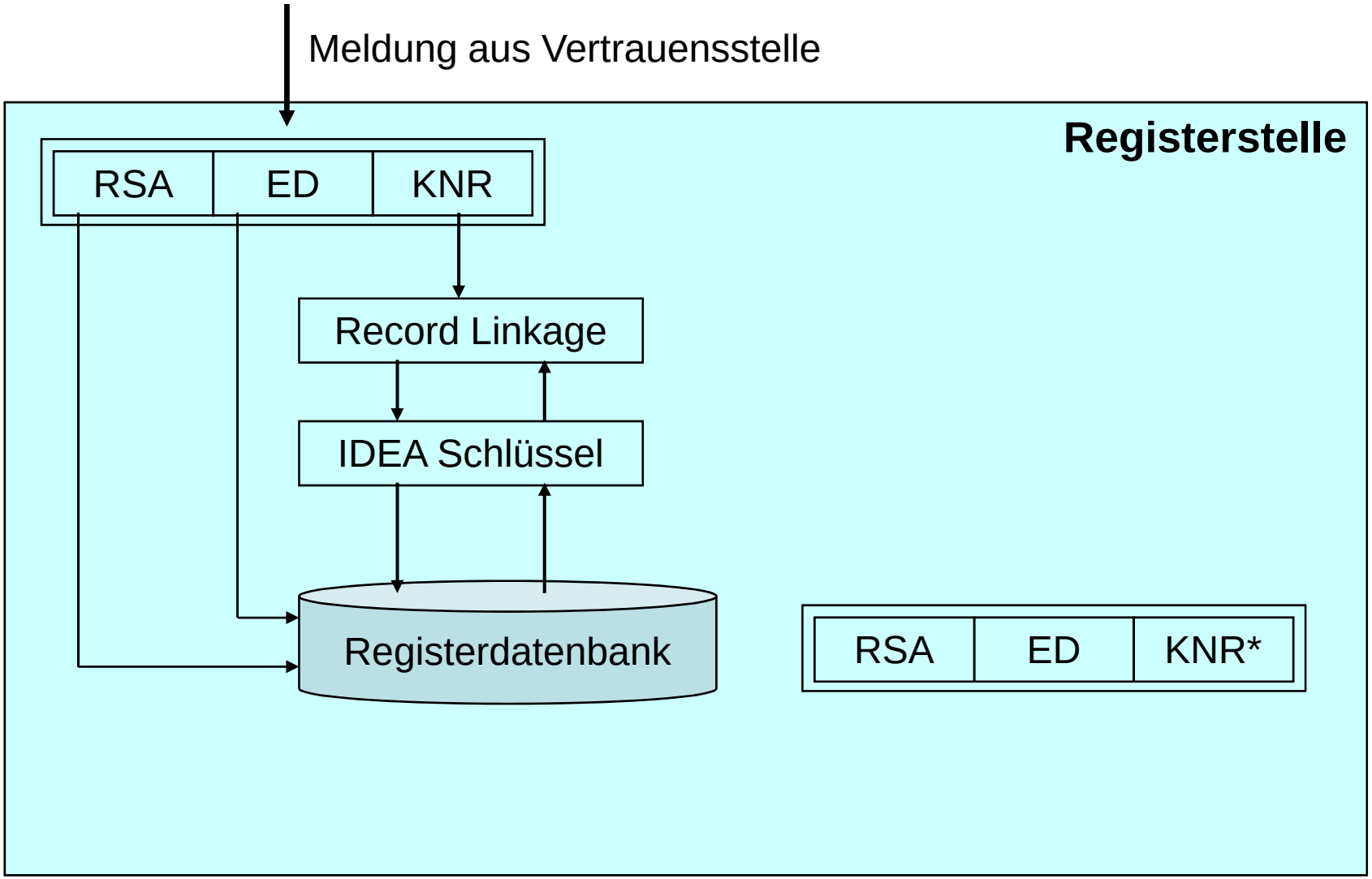
Das Ergebnis sind 22 Kontrollnummern zu je 23 Zeichen (ASCII)
Beispiel: HfY:8q#L;ANT+J+m$KR'x16

Die so in der Vertrauensstelle generierten Kontrollnummern lassen keine Identifizierung von Personen zu.

Pseudonymisierungslösungen im Krebsregister



Pseudonymisierungslösungen im Krebsregister



UNICON

Das Bundeskrebsregistergesetz sprach sich für ein einheitliches Verfahren zur Generierung von Kontrollnummern aus.
Die Entwicklung und Umsetzung erfolgte im Rahmen des UNICON-Projektes (uniform control number generator) durch OFFIS in Oldenburg.

Alle bundesdeutschen Landeskrebsregister wurden mit UNICON zur Generierung von Kontrollnummern und IDEA-Schlüsseln ausgestattet.

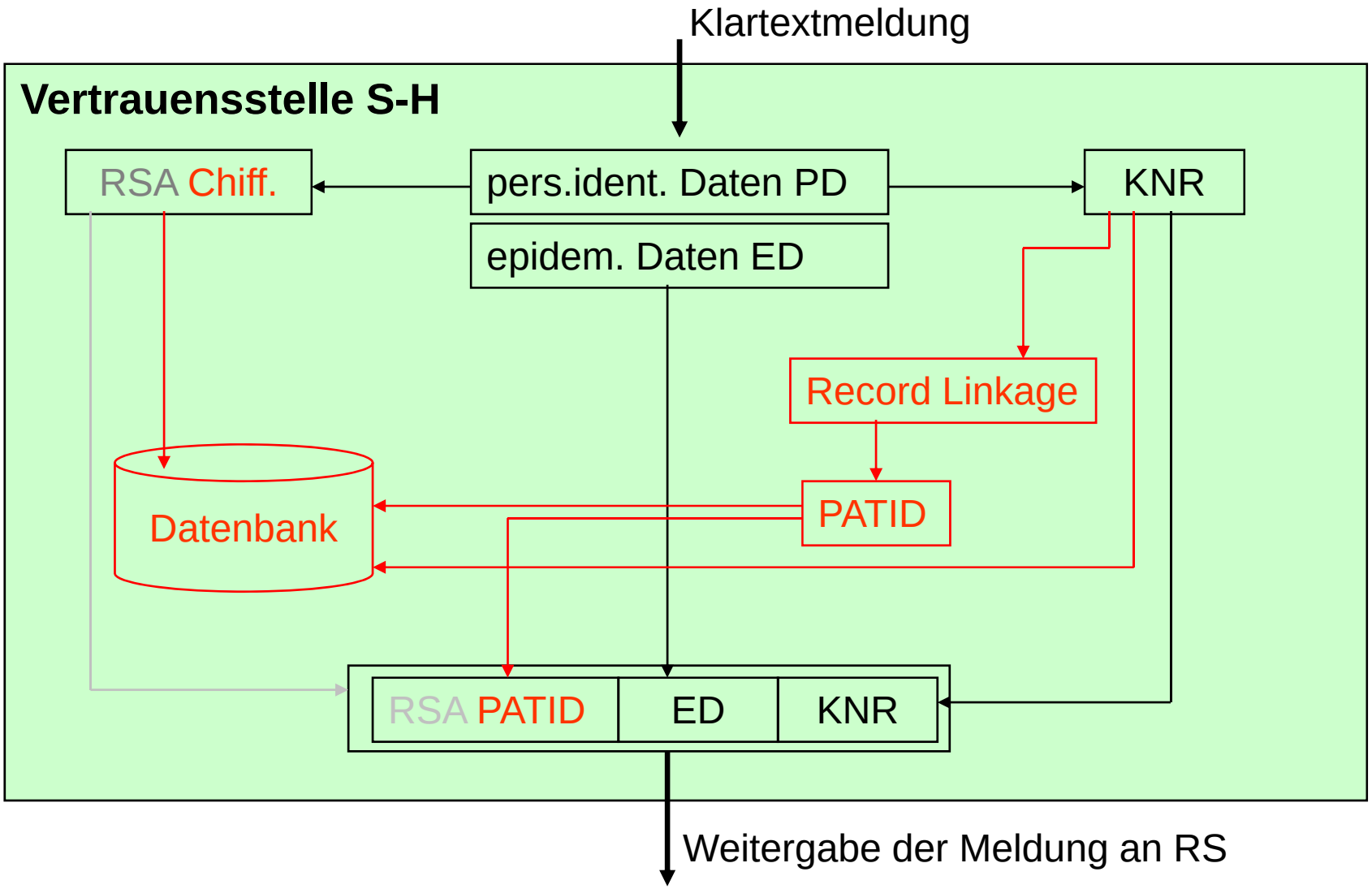
Das gesamte Konzept der Kontrollnummerngenerierung ist aus sicherheitstechnischer Sicht durch das Bundesamt für Sicherheit in der Informatik (BSI) akzeptiert worden.

Pseudonymisierungslösungen im Krebsregister

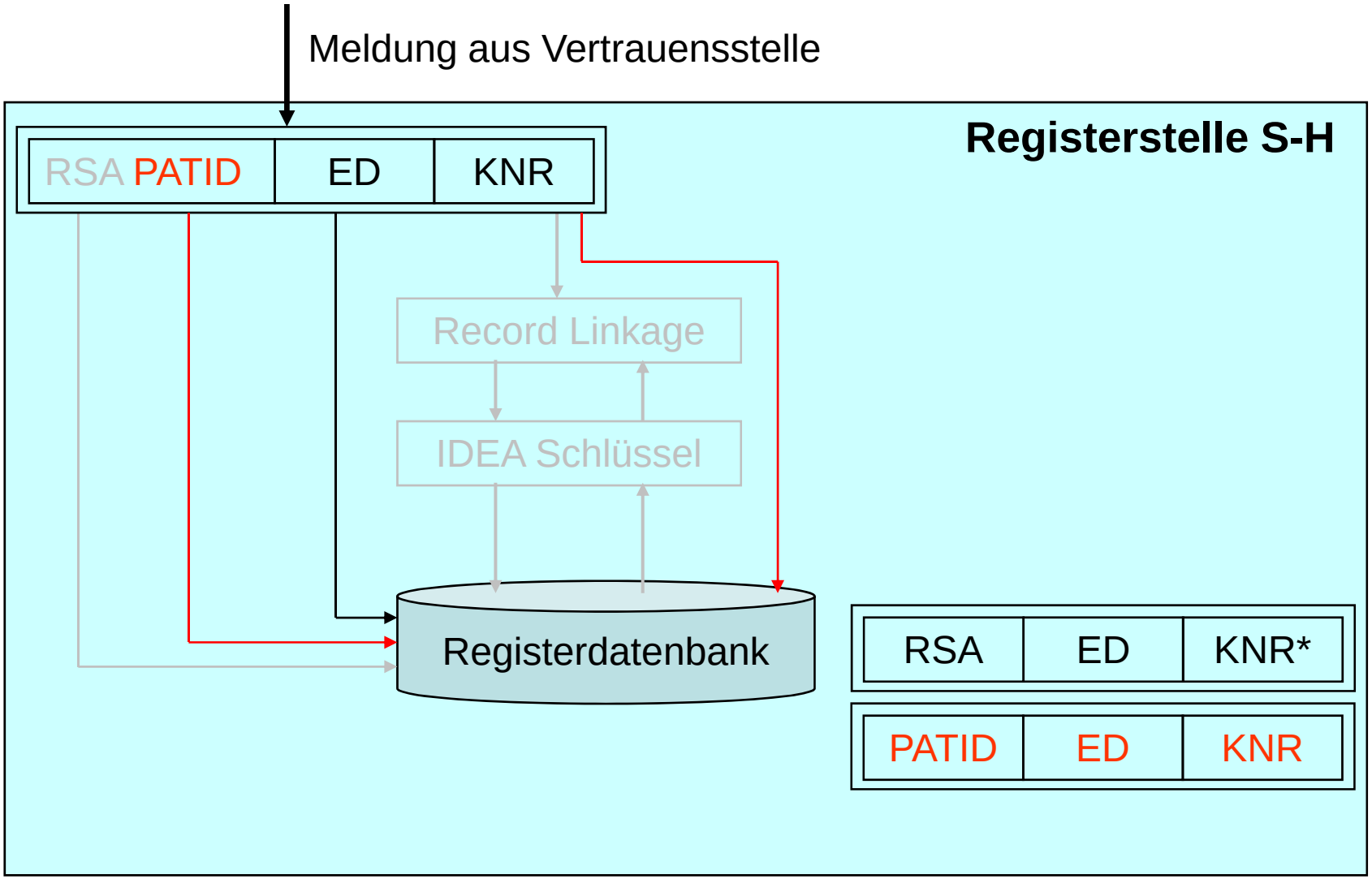
Pseudonymisierungslösungen im KR Schleswig-Holstein

- Räumliche Trennung von Vertrauens- und Registerstelle (Ärztekammer in Bad Segeberg / Institut für Krebsepidemiologie in Lübeck), damit verbunden auch die räumliche Trennung der Datenspeicherung: personenbezogene Daten in der Vertrauensstelle, epidemiologische Daten in der Registerstelle
- Vertrauensstelle: Entgegennahme und Bearbeitung der Meldungen, Trennung in Personen- und epidemiologische Daten, Bildung der Kontrollnummern, Verschlüsselung der personenidentifizierenden Daten. Löschen aller Klartextdaten. Record Linkage und Speicherung der personenbezogenen Daten
- Registerstelle: Bearbeitung der anonymisierten epidemiologischen Daten. Speicherung der epidemiologischen Daten und Kontrollnummern (im Linkage Format)

Pseudonymisierungslösungen im Krebsregister



Pseudonymisierungslösungen im Krebsregister

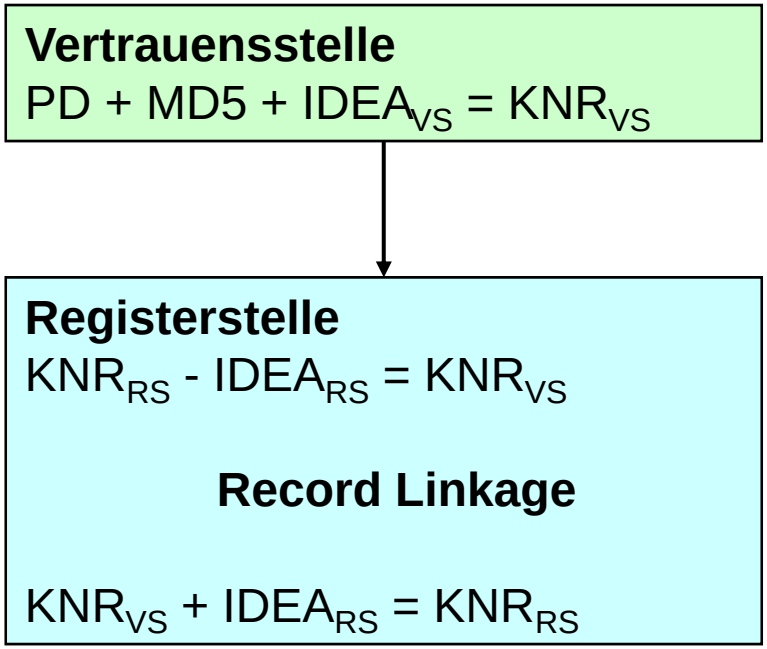


Anhang

Daten-Abgleiche auf Basis von Kontrollnummern

Innerhalb des Krebsregisters:

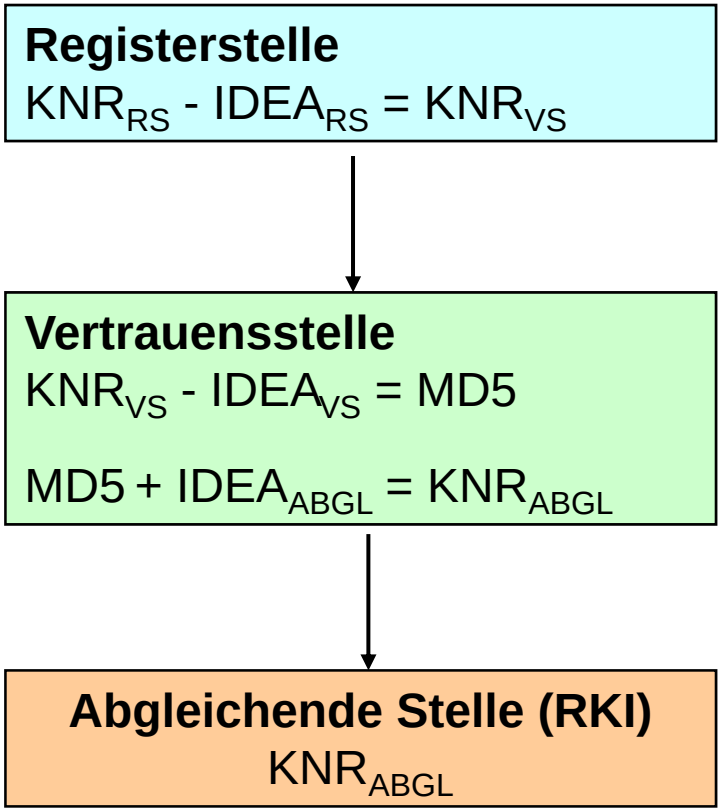
- In der Vertrauensstelle werden für neue Meldungen die Kontrollnummern erzeugt.
- Die Registerstelle stellt mit ihrem IDEA-Schlüssel die Kontrollnummern im Linkage Format wieder her.
- Durchführung des Abgleichs
- Bildung des Storage-Formats, löschen des Linkage-Formats



Daten-Abgleiche auf Basis von Kontrollnummern

Abgleich mit anderen Krebsregistern:

- Registerstelle: stellt mit ihrem IDEA-Schlüssel die Kontrollnummern für die abzugleichenden Daten im Linkage Format wieder her
- Vertrauensstelle macht ihre IDEA-Verschlüsselung rückgängig
- Durch Neuverschlüsselung des MD5-Chiffrats mit dem Abgleich-IDEA-Schlüssel werden die Kontrollnummern im Abgleichformat erstellt
- Weitergabe dieser Kontrollnummern an abgleichende Stelle, z.B. Robert Koch-Institut



Pseudonymisierungslösungen im Krebsregister

Zuordnung DDR-Namenskode

00 - Aa .. Am
01 - An .. Az
02 - Baa .. Bat
03 - Bau .. Beg
04 - Beh .. Ber
05 - Bes .. Bk
06 - Bl .. Bog
07 - Boh .. Bq
08 - Bra .. Brh
09 - Bri .. Bt
10 - Bu .. Bz
11 - C
12 - Da ..Dh
13 - Di .. Dq
14 - Dr .. Dz
15 - Ea.. Ell
16 - Elm .. Ez
17 - Fa.. Fh
18 - Fi .. Fj
19 - Fk .. Frh
20 - Fri.. Fz
21 - Ga .. Gek
22 - Gel ..Gln
23 - Glo .. Gq
24 - Gra .. Grn
25 - Gro .. Gz
26 - Haa .. Haj
27 - Hak .. Hase
28 - Hasf .. Heim
29 - Hein .. Heum
30 - Heun .. Hh
31 - Hi .. Hn
32 - Hoa .. Hofm
33 - Hofn .. Ht
34 - Hu .. Hz
35 - I
36 - Ja
37 - Jb .. Jz
38 - Kaa .. Kas
39 - Kat .. Kh
40 - Ki .. Kk
41 - Kla .. Klh
42 - Kli .. Kn
43 - Koa .. Kog
44 - Koh .. Kq
45 - Kra .. Krh
46 - Kri .. Kum
47 - Kun .. Kz
48 - La .. Ld
49 - Le .. Lh
50 - Li .. Log
51 - Loh .. Lz
52 - Maa .. Mar
53 - Mas .. Md
54 - Mea .. Mer
55 - Mes .. Miq
56 - Mir .. Muelleq
57 - Mueller .. Mz
58 - Na .. Nh
59 - Ni .. Nz
60 - O
61 - Pa .. Pe
62 - Pf .. Pk
63 - Pl .. Por
64 - Pos .. Pz
65 - Q
66 - Ra .. Reg
67 - Reh .. Rh
68 - Ri .. Rn
69 - Roa .. Ros
70 - Rot .. Rz
71 - Sa .. Scg
72 - Scha .. Schaq
73 - Schar .. Schj
74 - Schk .. Schmids
75 - Schmidt.. Schmz
76 - Schn .. Schq
77 - Schr .. Scht
78 - Schua .. Schul
79 - Schum .. Schz
80 - Sci .. Sh
81 - Si .. Sj
82 - Sk .. Ss
83 - Sta .. Stek
84 - Stel .. Stor
85 - Stos .. Sz
86 - Ta .. Th
87 - Ti .. Tz
88 - U
89 - V
90 - Wa .. Wd
91 - Wea .. Weim
92 - Wein .. Werl
93 - Werm .. Wik
94 - Wil .. Wn
95 - Wo .. Wz
96 - X
97 - Y
98 - Z
99 - Keine Angabe

Beispiel: Angela Mustermann, geb. Hansen => 570127 (je 2 Stellen vom Nach-, Vor- und Geburtsnamen)

Pseudonymisierungslösungen im Krebsregister

Kölner Phonetik

Die Kölner Phonetik arbeitet nach folgendem Algorithmus:
1. Leerzeichen, Bindestriche und ähnliche Sonderzeichen aus dem Namen entfernen.
2. Von mehrfach auftretenden Buchstaben nur den ersten beibehalten.
3. Umlaute und ß in ae, oe, ue und ss umsetzen.
4. Verbleibende Buchstaben nach folgendem Schema umsetzen:
Falls die nächsten beiden noch nicht bearbeiteten Buchstaben in der Diphtongtabelle (Diphtong = Doppellaut) vorhanden sind, dann dieses Buchstabenpaar ersetzen, anderenfalls nur den nächsten Buchstaben nach der Tabelle der Einzelbuchstaben ersetzen.