

# Konzept der CCP-IT des DKTK

Version 1.4.1 vom 30.01.2014

## Über dieses Dokument

Dieses Dokument beschreibt die IT der Clinical Communication Platform (CCP-IT) des Deutschen Konsortiums für Translationale Krebsforschung (DKTK), eines der vier BMBF-geförderten „Deutschen Zentren für Gesundheitsforschung“. Beabsichtigte Zielgruppen sind neben der Entwicklerrunde auch Verantwortliche für Tumordokumentationen und Biomaterialbanken am Standort sowie zuständige Datenschutzbeauftragte.

Das Dokument ist wie folgt strukturiert:

- Eine Kurzzusammenfassung (Seite 2) gibt einen schnellen Überblick über Zweck und Methodik der CCP-IT.
- Ein umfassendes Schaubild (Seite 3) stellt die zu entwickelnden Komponenten der CCP-IT und ihre Verantwortlichkeiten dar.
- Eine Komponentenreferenz (Seite 4) erläutert das Schaubild und zeigt für jede Komponente die beabsichtigte Funktionalität auf.
- Als ein „roter Faden“ durch die CCP-IT können die bereits diskutierten Datenflüsse und Anwendungsfälle (Seite 11) dienen.
- Die häufigsten im Rahmen der zukünftigen Diskussion aufkommenden Fragen werden im Listenformat als Fragen und Antworten (Seite 17) zusammengestellt werden.

Dieses Dokument ist noch unvollständig und beschreibt noch nicht die komplette CCP-IT in allen Details. Es stellt jedoch, wo bereits ausformuliert, den Konsens seiner Autoren dar.

## Autoren

Dieses Konzept wurde erstellt von M. Lablans<sup>1</sup> unter Mitwirkung von A. Borg<sup>1</sup>, J. Eils<sup>2</sup>, B. Felder<sup>2</sup>, E. Herzog<sup>2</sup>, D. Kadioglu<sup>1</sup>, F. Kärcher<sup>2</sup>, M. Koca<sup>3</sup>, M. Steffens<sup>1</sup> und F. Ückert<sup>1</sup>.

<sup>1</sup> Universitätsmedizin der Johannes Gutenberg-Universität Mainz

<sup>2</sup> Deutsches Krebsforschungszentrum, Heidelberg

<sup>3</sup> Universitäres Centrum für Tumorerkrankungen (UCT), Frankfurt

## Versionsgeschichte

Version	Datum	Änderung/Status	Autor
0.995	15.11.2012	Erste Fassung: Schaubild, Komponentenreferenz, Datenflüsse und Anwendungsfälle.	Martin Lablans
1.0	05.12.2012	Abkürzungsverzeichnis, Korrektur Abbildung Pseudonymisierung, Konsens durch AG CCP-IT	Martin Lablans, Sylvia Villain, Bernd Löffler
1.1	04.02.2013	Präzisierung BMB-Verwaltung und Studienregister, Hinweis auf Aufwand mehrerer Brückenköpfe, sprachliche Überarbeitung, viele Erklärungen auch für DKTK-Externe, Umstrukturierung	Martin Lablans
1.2	06.02.2013	Erklärungen für Datenflüsse und Anwendungsfälle	Martin Lablans
1.3	18.10.2013	Zusammenfassung, Anpassung an DS-Konzept, Anwendungsfälle, Kontrollnummernerzeugung	Martin Lablans, Andreas Borg

<b>1.4</b>	29.01.2014	Ergänzung um Pseudonymtrennung (DKTK#ID, MDS-ID), sprachliche und inhaltliche Präzisierung	Andreas Borg
<b>1.4.1</b>	30.01.2014	Korrektur in Übersichtsgrafik	Andreas Borg

## Kurzzusammenfassung

Die CCP-IT stellt die übergreifende Kommunikationsplattform des DKTK dar, die alle Standorte des Konsortiums vernetzt. Daten aus Tumordokumentationen und Biomaterialbanken können unter Bewahrung von Datenschutz und -hoheit für das Konsortium eingebracht werden, dank föderierter Netzwerkarchitektur auch von Prä-DKTK-Patienten. So werden Machbarkeitsschätzung und Rekrutierung von Studien nicht nur während, sondern schon vor deren Beantragung im DKTK unterstützt.

### 1.1 Komponenten der CCP-IT

An jedem Standort wird ein Brückenkopf betrieben. In diesen werden Patienten- und Probanddaten aus den lokalen Systemen überspielt. Der Brückenkopf ist zwar Software des DKTK, wird jedoch unter lokaler Hoheit betrieben; durch den Import verlassen noch keine Daten das Haus.

Ein Metadatenverzeichnis (MDR) stellt das gemeinsame Verständnis über alle im DKTK verwendeten Begriffe bereit. So sprechen alle Brückenköpfe und zentrale Komponenten in derselben Sprache. Auch Änderungen (etwa neue Definitionen aus entitätsspezifischen Task Forces) können schneller umgesetzt werden.

Eine zentrale MDS-Datenbank speichert eine Teilmenge der im MDR definierten Datensätze, die sog. Meldedatensätze (MDS). Diese werden vom Teiler des Brückenkopfs regelmäßig übermittelt. Die Suche in der zentralen MDS-Datenbank liefert schnell Ergebnisse und eignet sich beispielsweise für die Hypothesenbildung und Machbarkeitsprüfung für neue Studien.

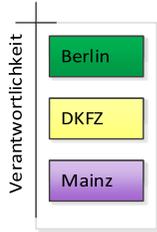
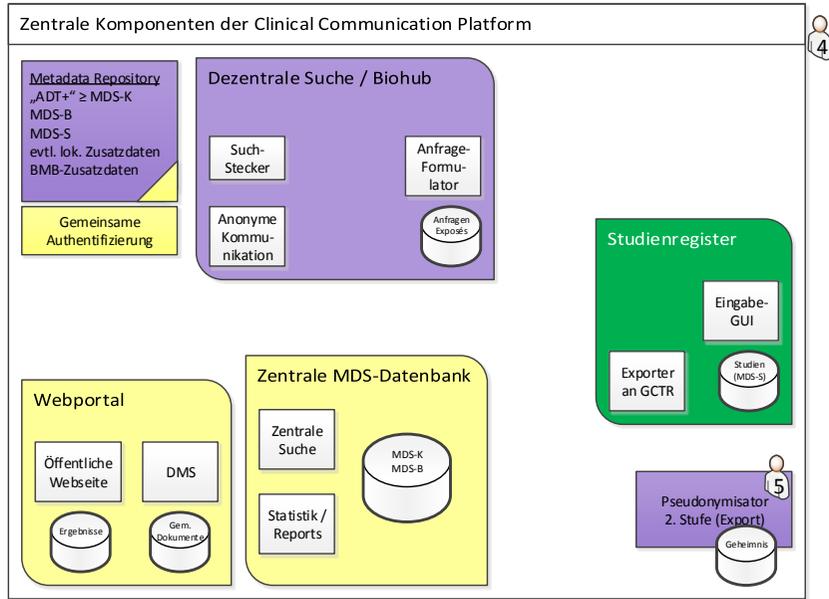
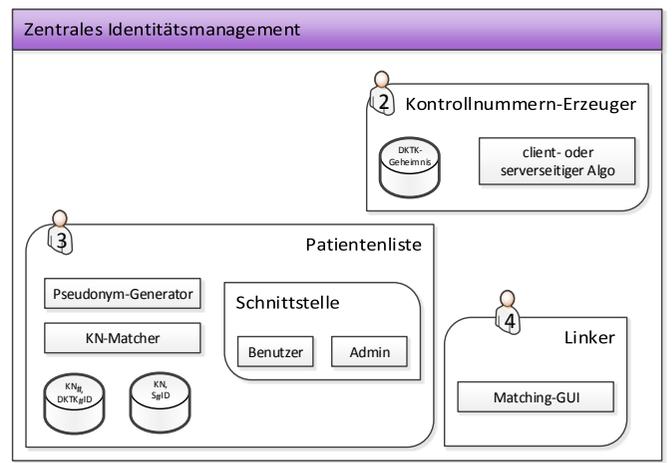
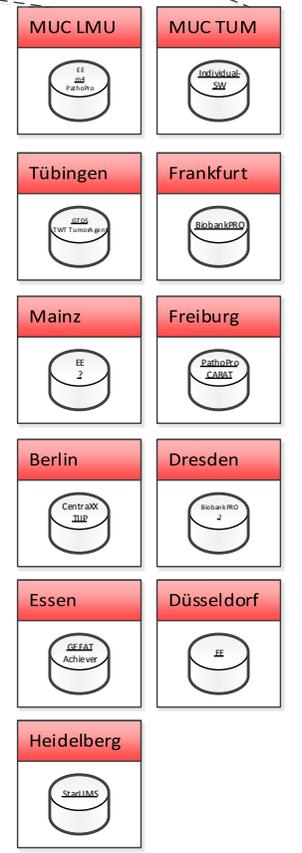
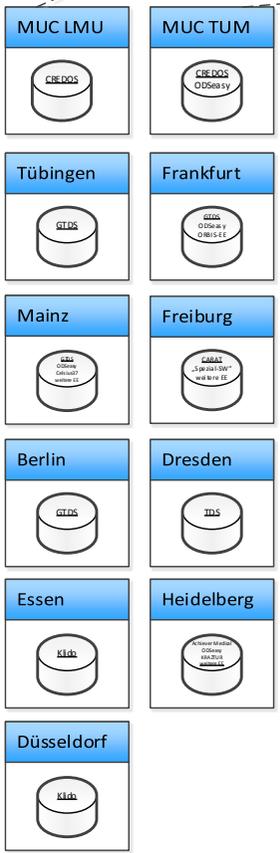
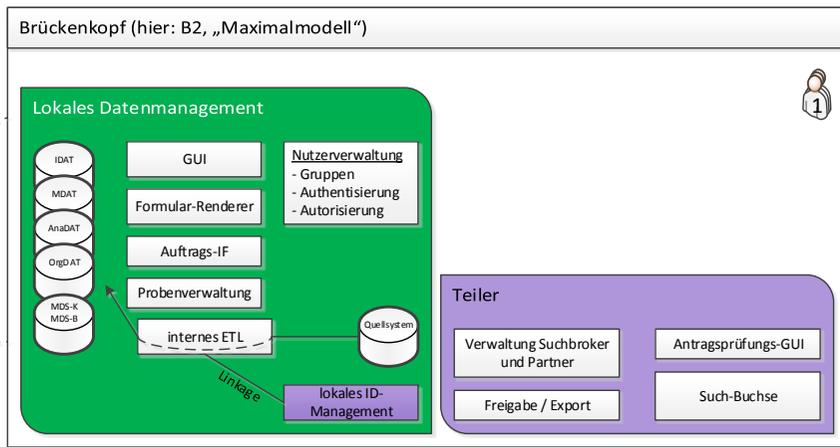
Durch die dezentrale Suche können auch Daten von Patienten angefragt werden, die vor Entstehung des DKTK lokal erfasst wurden. Das funktioniert auch ohne Bekanntgabe der Daten an das DKTK. Der Standort entscheidet hierbei selbst in einem manuellen Vorgang, welche Daten sein Haus verlassen – so kann er etwa lokale Entscheidungsprozesse oder die Patienteneinwilligung voll berücksichtigen. Die Suche ist dadurch langsamer, aber wesentlich gründlicher und eignet sich für die konkrete Rekrutierung von Probanden oder Material.

Das Identitätsmanagement vergibt Pseudonyme für Patienten. Patienten, die in die Verwendung Ihrer Daten im DKTK eingewilligt haben, können so auch nach Standortwechsel wiedererkannt werden.

Ein Studienregister führt Buch über geplante, beantragte und stattfindende DKTK-Studien. Ein- und Ausschlusskriterien können mithilfe der MDR-Elemente definiert werden.

## Die Arbeitsgruppe CCP-IT

Seit Ende 2011 besteht die Arbeitsgruppe CCP-IT. Sie setzt sich aus mindestens einem Vertreter jedes Standorts zusammen und entwickelt die IT der Clinical Communication Platform vom Konzept bis hin zur Implementierung.



## Komponentenreferenz

Die CCP-IT besteht aus zentralen und dezentralen Komponenten. Zentrale Komponenten sind das „zentrale Identitätsmanagement“ sowie die „zentralen Komponenten der Clinical Communication Platform“. Zentrale Dienste werden im DKTK genau einmal bereitgestellt (aber nicht notwendigerweise alle an derselben Stelle).

Dezentral laufen die sogenannten „Brückenköpfe“, die zwar im Rahmen der CCP-IT entwickelt werden, aber lokal an den Standorten<sup>1</sup> betrieben werden. Die Brückenköpfe werden in Zusammenarbeit mit den dortigen Verantwortlichen installiert, um IT-gestützt mit der CCP-IT zu kommunizieren.

Im Folgenden werden die Komponenten (Überschriften „N.“) und Module (Überschriften „N.M.“) der CCP-IT und Ihre Zusammenarbeit beschrieben. In Fußnoten ausgelagert sind Anmerkungen, die sich v.a. an Mitglieder der AG CCP-IT<sup>2</sup> richten.

## 2. Brückenkopf

Der Brückenkopf dient dazu, die Daten eines Standorts in ein DKTK-kompatibles Format zu überführen und für die anderen Komponenten nutzbar zu machen. Seine Aufgaben sind:

- Datenharmonisierung: Daten werden im Brückenkopf so harmonisiert und dupliziert gespeichert, dass sie von den übrigen Komponenten der CCP-IT verstanden werden.
- Sichtbarmachung für das DKTK: z.B. für zentrale und dezentrale Suche.
- Einhaltung von Datenhoheit: Der Brückenkopf erlaubt dem Standort eine Teilnahme an der CCP-IT auch ohne „Upload auf Verdacht“ seiner patientenbezogenen Daten an eine externe Stelle, was Datenschutz und Datenhoheit fördert.

Damit ist ein Brückenkopf jede im Rahmen der CCP-IT lokal in der Klinik installierte Software. Die Ausgestaltung ist aber standortspezifisch. Dabei werden die folgenden Szenarien unterschieden:

Standortstrategie Identifikation	Zusammenführung von klinischen und BMB <sup>3</sup> -Daten ist am Standort möglich	Klinische und BMB-Daten dürfen am Standort nicht zusammengeführt werden
Gem. Identifikator	<b>A1</b>	<b>B1</b>
IDAT an beiden Datenquellen	<b>A2</b>	<b>B2</b>
Kein Link vorhanden	<b>A3</b>	<b>B3</b>

Abbildung 1 - Szenarien Brückenkopfinstallation

<sup>1</sup> „Standort“ ist identisch zu „Zentrum“, es sei denn ein Zentrum ist auf mehrere Großinstitutionen aufgeteilt. In diesem Fall ist jede davon ein Standort. Beispiel: Das Zentrum „Frankfurt/Mainz“ besteht aus den Standorten Frankfurt und Mainz.

<sup>2</sup> In der Arbeitsgruppe CCP-IT leisten die von den jeweiligen Koordinatoren benannten Vertreter jedes DKTK-Standorts die Definition und Umsetzung der CCP-IT.

<sup>3</sup> Biomaterialbank

Erläuterung Achse „Verwaltung“ (A/B):

- Modell A: Am Standort besteht grundsätzlich die Möglichkeit, Daten aus Tumordokumentation und BMB-Verwaltung zusammen zu führen. Sei es, dass bereits ein System existiert, das beide Arten von Daten gemeinsam verwaltet oder aber es ist lokal möglich und gewollt diese zusammenzuführen.
- Modell B: Tumordokumentation und BMB-Verwaltung sind getrennt. Ein computergestütztes Mapping, gefolgt von einer lokalen Zusammenführung als Voraussetzung zum Verbinden von Daten aus Tumordokumentation und BMB-Verwaltung ist ausgeschlossen.

Erläuterung Achse „Identifikation“ (1/2/3):

- Modell 1: Es existiert ein gemeinsamer eindeutiger Identifikator, der einen Datensatz der Tumordokumentation mit seinem Pendant in der BMB-Verwaltung verknüpft. Für jedes Paar von Patienten gilt Gleichheit gdw. ihre Identifikatoren gleich sind.
- Modell 2: Es existiert zwar kein auf Gleichheit prüfbarer gemeinsamer eindeutiger Identifikator, aber es existieren sowohl auf Seiten der Tumordokumentation als auch der BMB-Verwaltung identifizierende Daten, über die ein fehlertolerantes Mapping (auch Record Linkage genannt) durchgeführt werden kann.
- Modell 3: Es existiert weder ein gemeinsamer Identifikator noch Record Linkage-taugliche Daten. Es gibt also keinen zuverlässig nutzbaren Link zwischen Tumordokumentation und BMB-Verwaltung.

Hier wird nicht unterschieden, ob die Zusammenführung lokal schon erfolgt oder noch zu leisten ist: Falls ein gemeinsamer Identifikator existiert (Fall 1) oder über den Zwischenschritt des Record Linkage hergestellt werden kann (Fall 2), geht das Modell davon aus, dass eine Zusammenführung der Daten aus Tumordokumentation und BMB-Verwaltung (im ETL-Prozess<sup>4</sup>) technisch möglich ist. Aus Sicht der zentralen Komponenten der CCP-IT sind diese Fälle gleich, denn in beiden Fällen kommunizieren die zentralen Komponenten der CCP-IT nachher mit einem einzigen Teiler (welcher auf beiden Datenbasen arbeitet). Anders sieht es im Fall B aus, in dem die Daten nicht einmal dann zusammengeführt werden dürfen, wenn es technisch möglich wäre. Dann müssen mehrere Brückenköpfe an einem Standort bereitstehen, von dem jeder einen der zu durchsuchenden Datenbasen teilt. Das ist z.B. der Fall, wenn datenschutzrechtlich lokal keine Genehmigung für eine solche Zusammenlegung existiert.

Es kann also, abhängig von lokalen Begebenheiten an einem Standort, mehrere Brückenköpfe geben, die jeweils für andere Patienten/Proben, Datenelemente oder Probandenkollektive zuständig sind. Für den Forscher wird dieser Umstand durch die CCP-IT weitgehend transparent gemacht, d.h. die CCP-IT gewährleistet ihm eine einheitliche Kommunikation auch mit komplexen Brückenkopf-Konstellationen.

## 2.1 Lokales Datenmanagement (Berlin, Mainz)

Bereitet die in den lokalen Primärsystemen prinzipiell vorliegenden, aber unterschiedlich strukturierten Datenbestände für eine Nutzung im DKTK auf. Bietet für Biomaterialbanken an den Standorten eine rudimentäre Proben- und Materialverwaltung samt Nutzer- und Gruppenverwaltung<sup>5</sup> und Formularhandling. Erhoben werden klinische Daten nach dem einheitlichen onkologischen Basisdatensatz (<http://www.tumorzentren.de/onkol-basisdatensatz.html>) sowie Daten zu Biomaterialproben.

Das System kann entweder verwendet werden um lokal manuell einzeln Daten zu erfassen (etwa für BMB, die so erstmals eine professionelle lokale Probenverwaltungssoftware bekommen) oder aber ein ETL-Prozess überführt

<sup>4</sup> Extraktion von Daten aus Quellsystemen, Transformation in ein Zielformat, Laden in ein Zielsystem, also Vereinigung von Daten aus ggfls. mehreren, unterschiedlich strukturierten Datenquellen in eine Zieldatenbank.

<sup>5</sup> Authentifizierung durch zentralen Authentifizierungsdienst der CCP wünschenswert?

Daten der bestehenden Primärsysteme in die Persistenz des lokalen Datenmanagements. Betrachtet man diese Attribute als Schlüssel-Wert-Paare, so wird dabei jeder Wert mit einem Schlüssel aus dem MDR versehen. Eine (pseudonymisierende) Zusammenführung von lokalen Datenquellen kann mithilfe des lokalen Identitätsmanagements (vgl. Abschnitt 2.3) erfolgen.

Mit jedem Standort wird – gemeinsam zumindest durch die Entwicklerstandorte Berlin und Mainz – einzeln abgestimmt, wie die klinischen Daten und die Biomaterialbankdaten in das lokale Datenmanagement gelangen können. Diese Abstimmungen sollen per Telekonferenz stattfinden. Ebenso sind bis zu drei Workshops für IT-Verantwortliche und Umsetzer aller Standorte geplant, bei denen solche Fragen spezifischer geklärt werden können. Es wird gemeinsam jedes System des jeweiligen Standortes unter Einbezug des lokal benannten Personals und dortigen Systemwissens oberflächlich analysiert. Im Falle von GTDS wird sogar der direkte Kontakt zum Hersteller durch die CCP-IT-Entwickler Berlin und Mainz aufgenommen, um den GTDS-nutzenden Standorten eventuell noch besser helfen zu können. Grundsätzlich sind lokale Implementationsarbeiten an den lokalen Systemen (außer dem Brückenkopf) Sache des jeweiligen Standortes, ungeachtet jeder angestrebten Arbeitserleichterung durch Verfügbarmachung und Anpassung unterschiedlicher Schnittstellen zum lokalen Brückenkopf.

## 2.2 Teiler (Mainz)

Leistet eine kontrollierte Freigabe der Datenbestände des lokalen Datenmanagements zur Nutzung durch Projekte des DKTK. Dabei kommen zwei sich ergänzende „Teilmethoden“ zum Einsatz: Eine zentrale Suche gibt sofort erste Ergebnisse aus, dann folgt eine langsamere, aber dafür umfassendere dezentrale Suche. Beide Suchen sind aber zentral initiiert, und zwar über zentrale Weboberflächen in der CCP-IT.

### Methode A: Zentrale Suche

Der Exporter lädt die Melddatensätze MDS-K und/oder MDS-B aller DKTK-Patienten<sup>6</sup> an die zentrale MDS-Datenbank hoch. Er berücksichtigt dabei etwaige DKTK-Flags, die im MDR definiert sind und Auskunft über die Auffindbarkeit und Verwendbarkeit des jeweiligen übermittelten Patienten/Probe im DKTK geben. Diese Daten können direkt von Forschern durchsucht werden (siehe Beschreibung der Zentralen Suche, Seite Zentrale Suche15).

### Methode B: Dezentrale Suche

Eine lokale Schnittstelle nimmt für den lokalen Standort Suchanfragen des Suchbrokers für die dezentrale Suche entgegen. Nach Konstruktion der iterativen Suche ist garantiert, dass der angesprochene Teiler für die angefragten Attribute zuständig ist. Er führt die Suche zunächst rein lokal durch und ermittelt einen Ergebnissatz. Ist dieser nichtleer, prüft er, ob der Anfragende als Forschungspartner bekannt und für automatische Anfragen autorisiert ist. Falls ja, wird der Ergebnissatz, ggfls. bzgl. des Datenschutzes aufbereitet<sup>7</sup>, sofort zurückgegeben. Falls nicht (das ist der Normalfall), wird sein Antrag, dargestellt als menschenlesbare Suchkriterien plus ein manuell vom Anfragenden erstelltes Projektexposé, dem lokal zuständigen Nutzer zusammen mit dem ermittelten Ergebnissatz vorgelegt. Der Nutzer bekommt daraufhin im lokalen Datenmanagement einen Hinweis, öffnet die Antragsprüfungsfläche und entscheidet darin über seine Antwort.

---

<sup>6</sup> Patient, für den eine DKTK-Einwilligungserklärung vorliegt.

<sup>7</sup> Details regelt das Datenschutzkonzept der CCP-IT.

## Schnittstelle zwischen lokalem Datenmanagement und Teiler

Der Teiler hält selbst nicht dauerhaft (Nutz-)Daten, sondern greift über eine definierte Schnittstelle auf den bereits aufbereiteten Bestand des lokalen Datenmanagements zu; nutzt dieses sozusagen als Persistenzschicht<sup>8</sup>. Eine zusätzliche Datenhaltung im Teiler entfällt. Der Teiler erfüllt folgende Aufgaben<sup>9</sup>:

- a) Er stellt eine Weboberfläche bereit, die im lokalen Datenmanagement als Menüpunkt eingebunden wird. Ein Tokensystem<sup>10</sup> gewährleistet Single-Sign-On<sup>11</sup> und übermittelt dem Teiler Sitzungsinformationen wie bspw. den bereits authentisierten Benutzernamen.
- b) Der Teiler meldet dem lokalen Datenmanagement, wenn seine Interaktion erforderlich ist (bspw. wenn eine Suchanfrage eingeht). So kann das Datenmanagement den Nutzer dazu veranlassen den Teiler aufzurufen, etwa durch eine E-Mail, ein Ticketsystem und/oder Hervorhebung des Menüpunkts.
- c) Der Teiler kann Suchanfragen auf dem lokalen Datenmanagement ausführen. Er übermittelt dazu Schlüssel-Wert-Paare, die die Eigenschaften der gesuchten Datensätze beschreiben. Das Datenmanagement antwortet mit einer Liste von Datensätzen. Diese enthalten mindestens einen Datensatzidentifikator, mit dem der Teiler sämtliche Eigenschaften abfragen kann. Dieser Identifikator bleibt für einen gegebenen Datensatz stets gleich.

## 2.3 Lokales Identitätsmanagement (Mainz)

Das lokale Identitätsmanagement stellt für die Pseudonymisierung sowohl von DKTK- als auch von nicht-DKTK-Patienten eine einheitliche Schnittstelle bereit. Es beinhaltet auch eine lokale Instanz des Kontrollnummernerzeugers (vgl. 3.1), mit dem Kontrollnummern für nicht-DKTK-Patienten erzeugt werden (vgl. Beschreibung der Pseudonymisierung in Abschnitt 2).

## 3. Zentrales Identitätsmanagement (Mainz)

Pseudonymisierung ist ein zur Aufrechterhaltung eines hohen Datenschutzniveaus notwendiger Schritt, um den Patienten vor Rückidentifizierung zu schützen. Anstelle seiner identifizierenden Daten (IDAT) treten Pseudonyme. In der CCP-IT kommen folgende Arten von Pseudonymen zum Einsatz:

- *S#ID*: Ein lokaler Identifikator, der nur für den Standort # eindeutig ist und keine standortübergreifende Verknüpfung von Daten eines Patienten erlaubt.
- *DKTK#ID*: Ein lokaler Identifikator, der nur für den Standort # eindeutig ist. Im Gegensatz zur *S#ID* können *DKTK#IDs*, die an verschiedenen Standorten zu einem Patienten existieren, in der zentralen Patientenliste einander zugeordnet werden.
- *MDS-ID*: Ein Identifikator, der nur innerhalb der zentralen MDS-Datenbank eindeutig ist. In der zentralen Patientenliste kann die *MDS-ID* den *DKTK#IDs* eines Patienten zugeordnet werden.

Im Falle einer geeigneten Einwilligung wird die jeweils standortspezifische *DKTK#ID*, erzeugt, welche auch (mithilfe der zentralen Patientenliste) eine Verfolgung des Patienten über Institutionsgrenzen hinweg erlaubt. Im Fall einer fehlenden DKTK-Einwilligung dürfen Klarnamen den Behandlungskontext des Standorts nicht verlassen; es wird dann ein lediglich die nur lokal vergleichbare *S#ID* generiert.

---

<sup>8</sup> Für den Nutzer wird zwar ersichtlich sein, dass es sich um zwei unterschiedliche Softwarekomponenten handelt. Durch ihre enge Verzahnung aber wird das den Arbeitsablauf nicht behindern. Es kann sogar von Vorteil sein, wenn der Benutzer leichter erkennt, wann er rein lokal mit seinen Daten arbeitet und wann er beginnt mit der Außenwelt zu kooperieren.

<sup>9</sup> Details sind in dieser frühen Projektphase noch nicht abschließend abgestimmt.

<sup>10</sup> Temporärer Zugriffsschlüssel, elektronisch ausgetauscht zwischen den Systemen, transparent für den Anwender.

<sup>11</sup> hier: Der Nutzer wird nicht erneut zu Eingabe eines Kennworts aufgefordert.

Das zentrale Identitätsmanagement ermöglicht eine datenschutzgerechte Zusammenführung („Record Linkage“) der von mehreren Standorten gesendeten patientenbezieharen Daten. Dazu werden drei Methoden/Werkzeuge kombiniert<sup>12</sup>, vgl. folgende Abschnitte 3.1-3.3 und Abbildung 2.

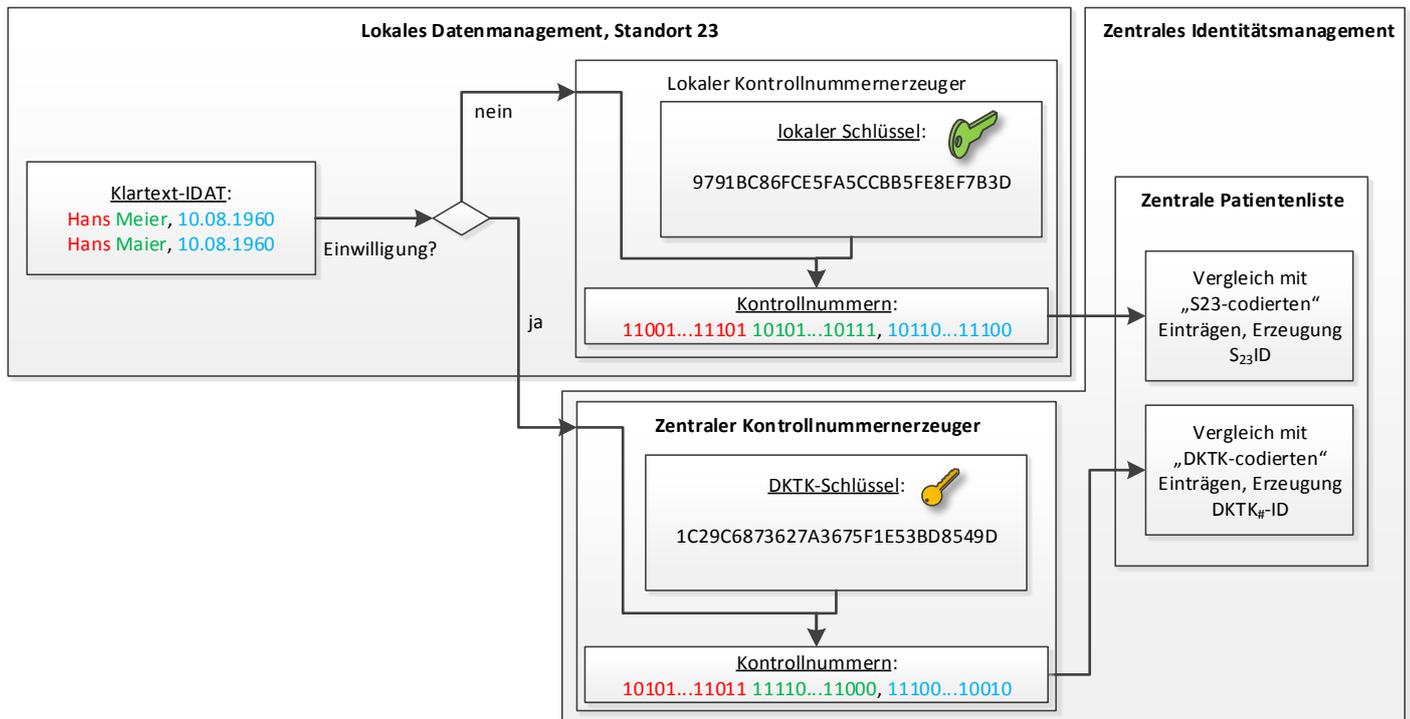


Abbildung 2 - Beispielhaftes Record Linkage-Verfahren. <sup>13</sup>

### 3.1 Kontrollnummern-Erzeuger

Wandelt Identifikationsdaten (Klartext-IDAT) durch eine spezielle Einwegverschlüsselung in unlesbare und nicht rückführbare, aber immer noch gewichtet vergleichbare Zeichenketten („Kontrollnummern“<sup>14</sup>) um. Um Wörterbuchattacken zu vermeiden, wird dabei ein Keyed-Hash Message Authentication Code verwendet, wobei der verwendete Schlüssel (im folgenden „Geheimnis“) für jede Instanz des Kontrollnummern-Erzeugers eindeutig und nur dort bekannt ist.<sup>15</sup>

Für die Pseudonymisierung von DKTK-Patienten kommt ein zentraler Kontrollnummern-Erzeuger zum Einsatz. Als weitere Sicherheitsmaßnahme (z.B. gegen Angriffe mit Probeverschlüsselungen), werden die dort erzeugten Kontrollnummern nicht an den Brückenkopf zurückgegeben, sondern direkt an die zentrale Patientenliste übermittelt. Die DKTK#ID wiederum wird nie dem zentralen Kontrollnummern-Erzeuger bekannt gemacht (Details siehe Abschnitt 2). Daraus folgt die Voraussetzung eines organisatorisch unabhängigen Betriebs des zentralen Kontrollnummernerzeugers.<sup>16</sup>

<sup>12</sup> Derzeitiges Maximalmodell. Ermittlung der genauen Anforderungen und Implementierung wird noch in einem Arbeitspaket geleistet.

<sup>13</sup> „Einwilligung? → nein“ bedeutet hier, dass noch keine DKTK-Einwilligung eingeholt wurde. Bei explizitem Widerruf gehen von vornherein keine Daten in die CCP-IT ein.

<sup>14</sup> Streng genommen sind das also keine Nummern. Idee und Name wurden (in angepasster Form) aus dem Bereich epidemiologischer Krebsregister übernommen.

<sup>15</sup> Für technische Details siehe: Schnell R, Bachteler T, Reiher J: Privacy-preserving record linkage using Bloom filters. BMC Medical Informatics and Decision Making 2009, 9:41. <http://www.biomedcentral.com/1472-6947/9/41>

<sup>16</sup> Das bedeutet: Betrieb ist noch zu klären. Mainz wird das aber „schlüsselfertig“ bereitstellen.

Bei nicht-DKTK-Patienten erfolgt die Kontrollnummernerzeugung im lokalen Identitätsmanagement.

### 3.2 Patientenliste

Vergibt für eine Gruppe von Kontrollnummern eines Patienten einen Personenidentifikator (DKTK#-ID, S#ID, MDS-ID). Ihr Kontrollnummern-Matcher kann dabei Patienten selbst bei abweichender Schreibweise und im Fall der DKTK-ID auch aus unterschiedlichen Standorten wiedererkennen.

### 3.3 Manuelles Linken

Eine Schnittstelle erlaubt einem Menschen, Ergebnisse des automatischen Matching zu überprüfen und ggfls. zu korrigieren, d.h. Duplikate zusammenzuführen oder fälschlicherweise zusammengeführte Datensätze zu trennen. Hierfür werden die Matchgewichte (Vergleichswerte zwischen den einzelnen Attributen von zu prüfenden Patienten) angezeigt und es besteht die Möglichkeit, zur Entscheidungsfindung auf medizinische Daten zuzugreifen. Diese Komponente ist deshalb beim Betreiber der zentralen MDS-Datenbank, wo ohnehin MDAT von DKTK-Patienten gespeichert sind, angesiedelt.

### 3.4 Informationelle Gewaltenteilung

Konsequent durchgeführt wird eine informationelle Gewaltenteilung, angedeutet durch die nummerierten Personenbilder (4). Das bedeutet, dass die Komponenten mit unterschiedlichen Nummern logisch, physikalisch und organisatorisch getrennt voneinander laufen, was die Gefahr eines Datenlecks verringert:

- Das zentrale Identitätsmanagement (2,3) wird getrennt betrieben von den übrigen zentralen Komponenten der CCP-IT (4). So kann jemand, der in der CCP-IT auf klinische oder BMB-Daten Zugriff hat, diese keinem realen Patienten zuordnen.
- Innerhalb des zentralen Identitätsmanagements wird der Kontrollnummernerzeuger (2; zu schützen ist hier das verbundweite Geheimnis) getrennt betrieben von der Patientenliste (3). So ist sichergestellt, dass die in der Patientenliste gespeicherten Kontrollnummern keinen direkten Rückschluss auf die Identität des Patienten zulassen.
- Selbiges gilt natürlich für lokale Geheimnisse.

## 4. Zentrale Komponenten der Clinical Communication Platform

Die CCP-IT besteht aus allen zentral laufenden Komponenten außer dem zentralen Identitätsmanagement.

### 4.1 Metadata Repository (Mainz, Heidelberg)

Schafft Vergleichbarkeit von an unterschiedlichen Orten erfassten Datenelementen. Dazu kennt es die Bedeutung (Semantik) sämtlicher im DKTK verwendeten (Nutz-)Datenelemente. Bietet als Folge insbesondere ein kontrolliertes Vokabular (Syntax), kann aber auch maschinenlesbare, strukturierte Aussagen über Datenelemente machen, bspw. konzeptuelle Domänen oder Wertebereiche. Wird initial durch Mainz mit den Meldedatensätzen des DKTK (MDS-K, MDS-B, MDS-S) sowie für die CCP-IT notwendige technische Attribute (Pseudonyme, „DKTK-Flag“, ...) gefüllt und kann später erweitert werden.

Das bedeutet: Für jedes Schlüssel-Wert-Paar<sup>17</sup>, das – ganz gleich ob zentral oder am Standort – im DKTK erfasst wird, steht hier der wohldefinierte Schlüssel.

---

<sup>17</sup> Beispiel: Schlüssel = „Blutdruck systolisch, Patient bei Messung aufrecht sitzend, in mmHg“, Wert = 125

## 4.2 Zentrale MDS-Datenbank (Heidelberg)

Nimmt mithilfe einer Webschnittstelle die von den Exportern der Teiler verschickten MDS-K und MDS-B entgegen und verwahrt sie in der zentralen MDS-Datenbank in Heidelberg. Die Speicherung erfolgt zusammen mit der MDS-ID, um die Zuordnung verschiedener Datensätze eines Patienten zueinander möglich zu machen. Autorisierte Nutzer können in einer Suchmaske mittels Suchkriterien aus MDS-K und MDS-B Anfragen auf diesem Datenbestand ausführen. Als Ergebnis wird die Anzahl passender Datensätze an jedem Standort zurückgegeben. Optional können gefundene Datensätze angezeigt und exportiert werden. Formulierten Suchanfragen können an den dezentralen Suchbroker (vgl. 4.3) übergeben werden. Statistiken und Reports auf den zentralen DKTK-Datenbeständen können hier generiert werden.

## 4.3 Suchbroker für die dezentrale Suche und Biomaterialhub (Mainz)

Frägt dezentral klinische Daten und Biomaterialdaten ab. Nachdem der anfragende Forscher seine gesuchten Patienten oder Proben formuliert hat, lädt er ein Projektexposé<sup>18</sup> hoch. Der Suchbroker verwendet dann ein iteratives Verfahren, das für jeden Standort zunächst ermittelt<sup>19</sup>, wo dieser die für die Suchanfrage relevanten Items ablegt. Ergebnis sind die Adressen der zuständigen Teiler unter Bezug auf die nachgefragten Datenelemente. Die Teiler werden dann nacheinander abgefragt und ihre zurückgegebenen Suchergebnisse kombiniert<sup>20</sup>.

## 4.4 Gemeinsamer Authentifizierungsdienst (Heidelberg)

Wird von anderen Komponenten verwendet, um Nutzer gegen ein im DKFZ betriebene zentrale Nutzerdatenbank (nicht Teil der CCP-IT) zu authentifizieren<sup>21</sup>.

## 4.5 Webportal (Heidelberg)

Stellt den öffentlichen Webauftritt der CCP-IT für das DKTK bereit. Es dient als einheitlicher Einstieg für alle zentralen Komponenten der CCP-IT. Bietet außerdem die Publikation von Forschungsergebnissen und einen nichtöffentlichen Bereich, der u.a. zum Austausch von internen Dokumenten verwendet wird.

## 4.6 Pseudonymisator 2. Stufe (Mainz)

Um eine Rückidentifizierung einer beschriebenen Person zu erschweren, ist es möglich, beim Export von personenbezogenen Daten an Forscher nicht eines der in der CCP-IT verwendeten Pseudonyme (DKTK<sub>#</sub>ID, MDS-ID, S<sub>#</sub>ID) herauszugeben, sondern ein für das vorliegende Forschungsprojekt eigenes, Follow-Up-fähiges Pseudonym zweiter Stufe (CCP-PSN-Projekt<sub>#</sub>) zu erzeugen.

## 4.7 Studienregister (Berlin)

Führt zentral nach einem MDS-S Beschreibungen (keine Verlaufsdocumentation) von geplanten, beantragten und stattfindenden DKTK-Studien, die manuell durch die einzelnen Standorte eingegeben werden. Dokumentationspunkte werden aus dem MDR bezogen, können also insbesondere erweitert werden. Ein- und Ausschlusskriterien können anhand des MDS-K definiert werden. Suchen und Standardreports werden manuell oder als Systemreport ausgegeben. Ein Studienadministrator kann die Pflege der Dokumentationspunkte mit Kontakten und Vokabeln zentral durchführen.

---

<sup>18</sup> Im einfachsten Fall ein manuell erstelltes PDF, das für das geplante Vorhaben wirbt. Denkbar wäre natürlich auch eine Verknüpfung mit der Study Registry.

<sup>19</sup> Sollte keine Probleme in Firewalls geben, denn ein Upload wird ja auch ermöglicht.

<sup>20</sup> Mit diesem iterativen Suchverfahren kann der zusätzliche „zentrale Brückenkopf pro Standort“, wie er zunächst in unserem Meeting in Frankfurt scheinbar nötig war, eingespart werden. So wird die minimal machbare Anzahl an Brückenköpfen erreicht: Einer pro Datenhalter.

<sup>21</sup> Kerberos oder ein föderiertes Authentifizierungsverfahren wie OpenID, ADFS, ...

# Datenflüsse und Anwendungsfälle

Im Folgenden werden die bereits diskutierten CCP-IT-Anwendungsfälle identifiziert und die dazu nötigen Datenflüsse modelliert. Die Reihenfolge ist so gewählt, dass die Anwendungsfälle aufeinander aufbauen. So entsteht ein „roter Faden“ zur Funktionsweise der CCP-IT.

## 1. Import in Brückenkopf<sup>22</sup>

Voraussetzung: Installierter Brückenkopf.

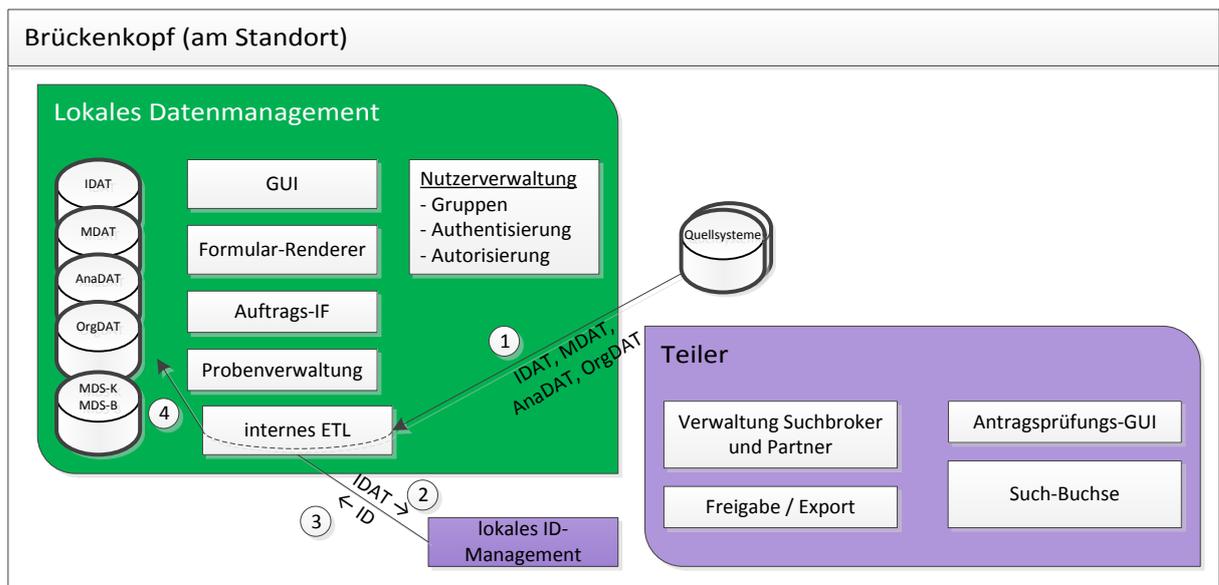


Abbildung 3 - Import von Daten in den Brückenkopf.<sup>23</sup>

Abbildung 3 zeigt, wie Daten aus Quellsystemen eines Standorts in den Brückenkopf importiert werden:

- 1) Identifizierende, medizinische und Probanddaten werden aus mehreren Quellsystemen durch einen ETL-Prozess extrahiert.
- 2) Identifizierende Daten werden mithilfe des lokalen Identitätsmanagements mit einem primären Patientenidentifikator erster Stufe versehen (Details siehe Anwendungsfall 2: Pseudonymisierung)
- 3) Der ETL-Prozess ordnet den Identifikator dem Datensatz zu.
- 4) Daten werden im lokalen Datenmanagement abgelegt.

Die Speicherung von identifizierenden Daten zusammen mit klinischen oder Probanddaten kann unkritisch sein, solange diese (wie im Brückenkopf) den Standort nicht verlassen. Sollte an einem Standort die lokale Zusammenführung von klinischen und Biobankdaten ausgeschlossen sein, können mehrere Brückenkopf-Instanzen zum Einsatz kommen, die dann unter getrennter lokaler Hoheit stehen. An einem Standort stehen dann mehrere Brückenköpfe, z.B. einer für die Daten der Tumordokumentation und einer für die der Biobank, vgl. Abbildung 4. Nach Möglichkeit sollte jedoch nur ein Brückenkopf pro Standort zum Einsatz kommen, um den Aufwand insbesondere für die lokalen Administratoren zu minimieren.

<sup>22</sup> To-Do: Abgleich mit Metadatenverzeichnis ergänzen.

<sup>23</sup> Beispiele: IDAT = Vorname, Nachname, Geburtsdatum, DKTK-Einwilligung; MDAT = klinische Daten (Obermenge des MDS-K); AnaDAT = Daten zur Bioprobe (Obermenge des MDS-B); OrgDAT = Verwaltungsdaten zur Probe

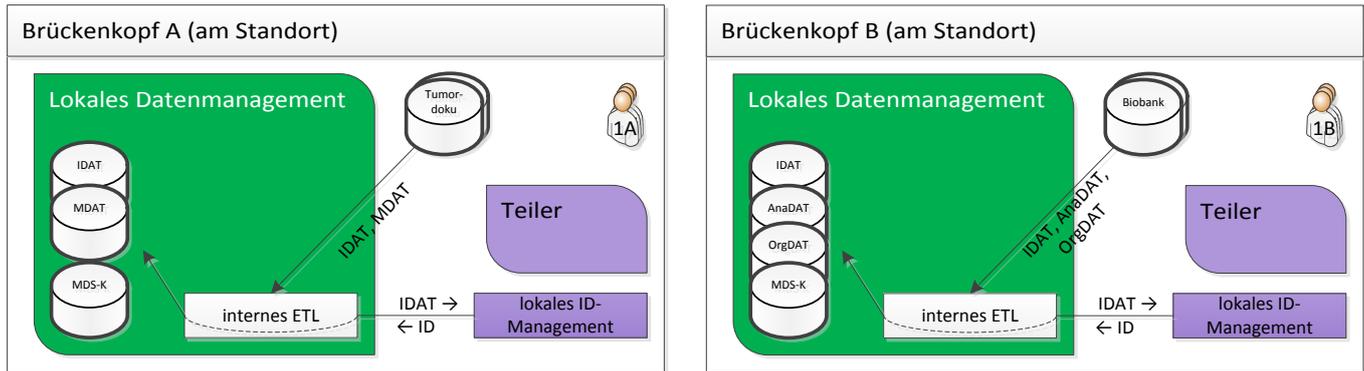


Abbildung 4 - Beispiel für Verwendung mehrerer Brückenköpfe an einem Standort. Brückenkopf A erhält nur die Daten der Tumordokumentation, Brückenkopf B nur die der Biobank. Beide werden von denselben Systembetreuern gewartet wie ihre Quellsysteme.

## 2. Pseudonymisierung

Benutzt durch: Anwendungsfall 1: Import in Brückenkopf.

Aus Datenschutzgründen entstehen zwei verschiedene Datenflüsse, je nachdem ob für den zu pseudonymisierenden Patienten eine DKTK-spezifische Patienteneinwilligung vorliegt („DKTK-Patient“) oder nicht. Die Fälle unterscheiden sich wie folgt:

- Fall mit Einwilligung: Es wird ein lokales, aber mithilfe der zentralen Patientenliste verbundweit verknüpfbares Pseudonym (DKTK#ID) erzeugt, das eine Zuordnung des Patienten auch über Standortgrenzen hinweg erlaubt.
- Fall ohne Einwilligung: Es dürfen keine IDAT den Standort verlassen. Es wird ein lokales Pseudonym (S#ID) erzeugt, d.h. eines, das nur innerhalb des Standorts zugeordnet werden kann. Man kann es etwa dazu verwenden um passendes Biomaterial zu einem klinischen Datensatz zu finden. Es ist aber nicht möglich, einen Patienten über Standortgrenzen hinweg zu verfolgen.

Abbildung 5 zeigt, wie die Pseudonymisierung im Rahmen des ETL-Prozesses genutzt wird um ein Pseudonym zu erhalten.

- 1) Das lokale Identitätsmanagement erhält die Klartext-IDAT eines Patienten.
- 2) Klartext-IDAT werden übertragen an den Kontrollnummern-Erzeuger...
  - a) im Fall mit Einwilligung: ...des zentralen Identitätsmanagements.
  - b) im Fall ohne Einwilligung: ...des lokalen Identitätsmanagements.
- 3) Der jeweilige empfangende Kontrollnummern-Erzeuger errechnet aus den Klartext-IDAT und seinem Geheimnis Kontrollnummern (KN: DKTK-weit vergleichbare Kontrollnummern, KN#: Nur für den Standort # vergleichbare Kontrollnummern).
- 4) nur im Fall mit Einwilligung: Der zentrale KN-Erzeuger übermittelt die (DKTK-weit vergleichbaren) Kontrollnummern an die Patientenliste. Diese erstellt ein temporäres Kontrollnummern-Ticket (KNTKT) und gibt dies an das lokale Identitätsmanagement zurück.
- 5) Das lokale Identitätsmanagement übermittelt die erzeugten Kontrollnummern, oder, im Fall mit Einwilligung, das eben erhaltene Ticket KNTKT, an die Patientenliste des zentralen Identitätsmanagements. Durch die Verwendung des Tickets bleiben die DKTK-weit vergleichbaren KN dem lokalen Identitätsmanagement verborgen.

- 6) Das zentrale Identitätsmanagement gleicht die erhaltenen Kontrollnummern mit den bestehenden ab (KN-Matcher) und gibt im Falle eines Treffers eine bestehende ID (DKTK#ID bzw. S#ID) zurück oder erzeugt eine neue mithilfe des Pseudonym-Generators.

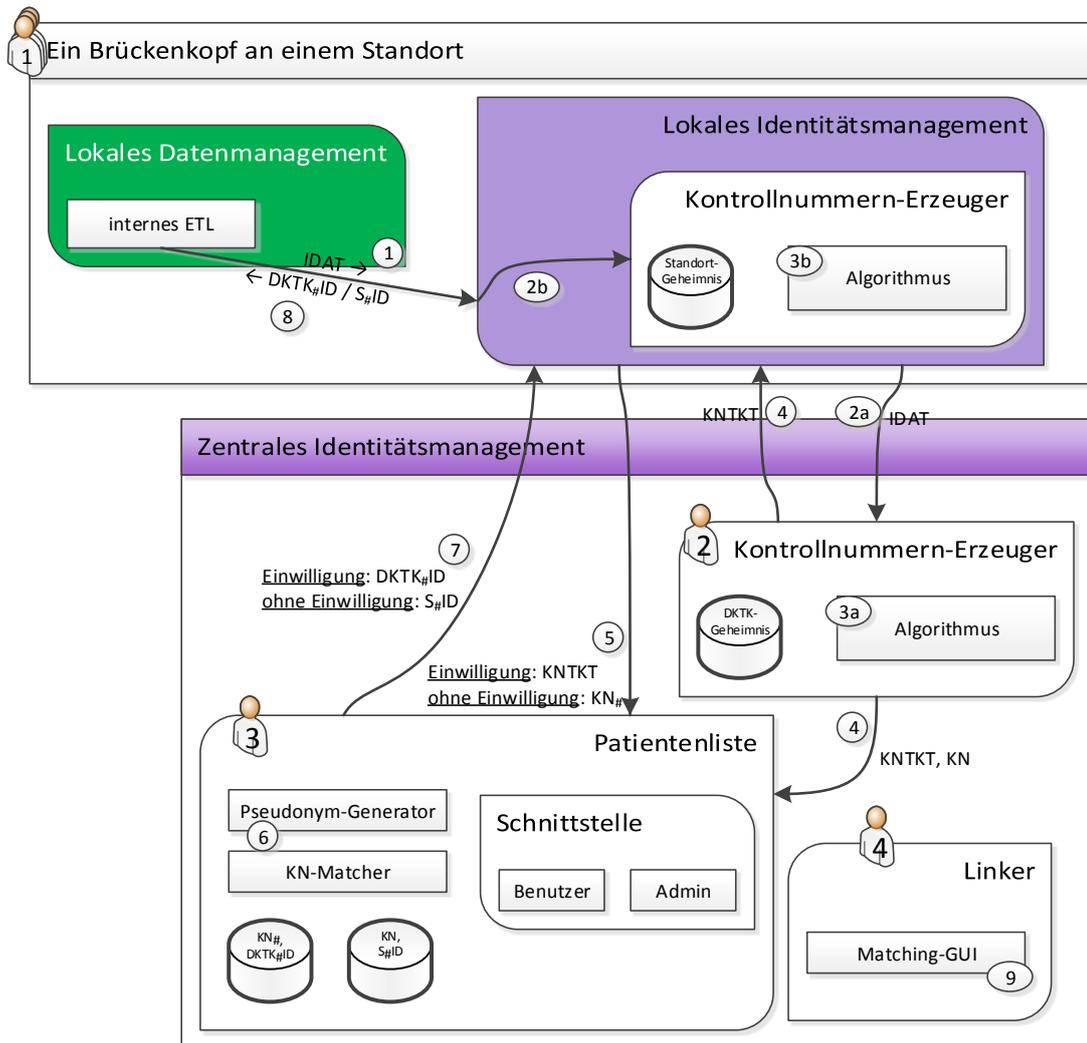


Abbildung 5 - Erzeugung lokaler und verbundweiter Pseudonyme.

- 7) Das lokale Identitätsmanagement erhält als Antwort auf seine Anfrage die ID. Es erfährt nicht, ob der Patient bereits bekannt war.
- 8) Die ID wird als Antwort auf die ursprüngliche Anfrage zurückgegeben. Es ist nicht erkennbar, ob der Patient bereits bekannt war. Es ist aber erkennbar, ob es sich um eine lokale S#ID oder um eine verbundweit verknüpfbare DKTK#ID handelt.
- 9) (optional) Sollte die Patientenliste sich nicht sicher sein, ob ein Patient schon in der Liste steht oder nicht, erfährt das der für das Record Linkage verantwortliche Administrator (4). Er entscheidet manuell auf Basis von Matchgewichten (aber keinen IDAT!) und – im Fall mit Einwilligung – Daten in der CCP-IT (v.a. MDAT), ob es sich um denselben Patienten handelt<sup>24</sup>.

<sup>24</sup> To-Do: Konzept für späteren Austausch/Wechsel eines Pseudonyms erster Stufe überlegen.

### 2.1 Erläuterung: Geheimnisse und Gültigkeit der Pseudonyme

Die Fälle mit Einwilligung und ohne Einwilligung unterscheiden sich im Geheimnis, das für die Kontrollnummernerzeugung verwendet wird: Das DKTK-weite Geheimnis, das genau dem Kontrollnummern-Erzeuger des zentralen Identitätsmanagements bekannt ist (Fall mit Einwilligung), erlaubt beim Matchen durch die zentrale Patientenliste eine Zuordnung des Patienten über Standortsgrenzen hinweg. Die lokalen Geheimnisse hingegen, die genau den Kontrollnummern-Erzeugern der lokalen Identitätsmanagement-Instanzen in allen Brückenköpfen desselben Standorts bekannt sind (Fall ohne Einwilligung), erlauben nur einen Abgleich der Patienten desselben Standorts. Das ist nötig um z.B. zu einem gegebenen klinischen Datensatz das passende Biomaterial zu finden. Lokale Kontrollnummern werden, obgleich sie nur lokal Sinn ergeben, durch die zentrale Patientenliste abgeglichen. Das erspart die lokale Installation einer Patientenliste, wahrt aber weiterhin den Datenschutz, da die Kontrollnummern per Konstruktion eine Reidentifikation außerhalb des Standorts nahezu unmöglich machen.

### 3. Upload in zentrale MDS-Datenbank

Der Upload in die zentrale MDS-Datenbank erfolgt automatisch nach einem festen Zeitintervall (in der Regel täglich). Dabei werden aktuelle (d.h. seit dem letzten Upload hinzugekommene) medizinische Daten von DKTK-Patienten gemäß der Meldedatensätze MDS-K und MDS-B vom Teiler aus dem lokalen Datenmanagement ausgelesen und über eine sichere HTTP-Verbindung in die zentrale MDS-Datenbank exportiert.

Um die richtige Zuordnung der Daten in der MDS-Datenbank zu gewährleisten, werden diese beim Export mit Hilfe der zentralen Patientenliste umpseudonymisiert, d.h. die DKTK#ID wird durch die zugehörige MDS-ID ersetzt. Dazu kommt ein asymmetrisches Verschlüsselungsverfahren zum Einsatz, wobei der private Schlüssel nur der zentralen MDS-Datenbank bekannt ist. Der Ablauf ist wie folgt (vgl. das Sequenzdiagramm in Abbildung 6):

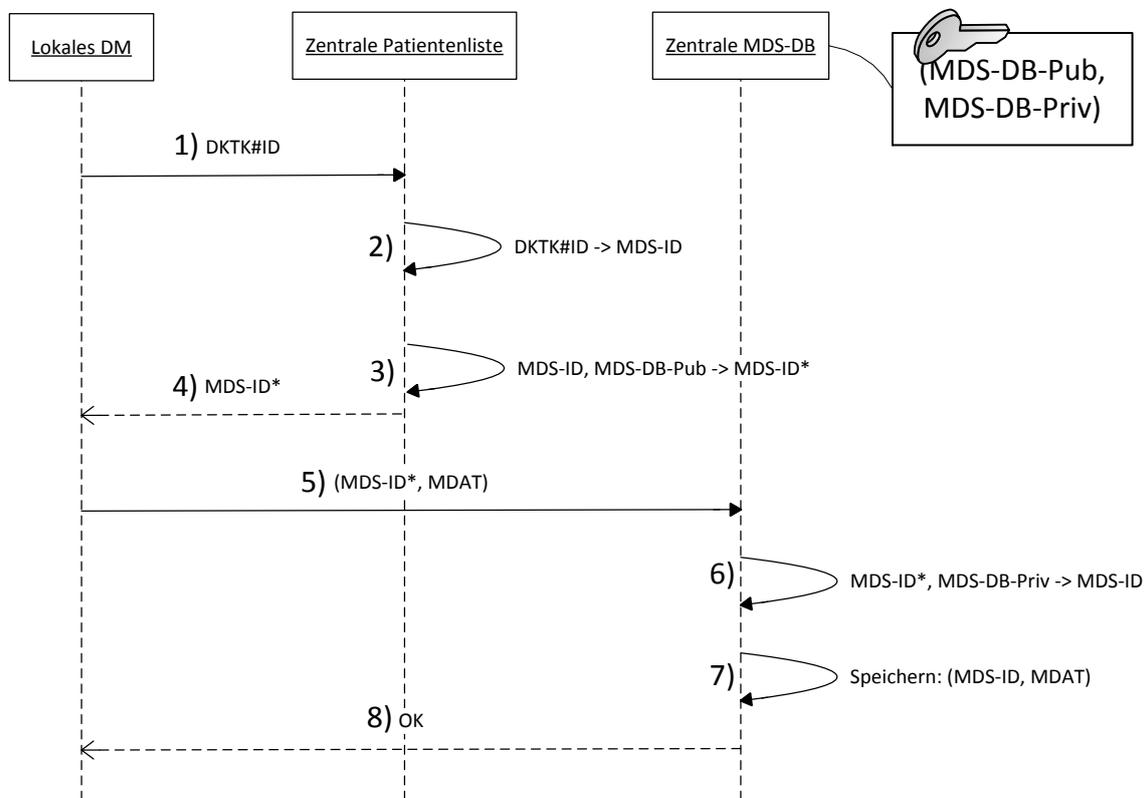


Abbildung 6 - Umpseudonymisierung beim Upload in die zentrale MDS-Datenbank

1. Der Teiler übermittelt die DKTK#ID des betreffenden Patienten an die zentrale Patientenliste.
2. Die zentrale Patientenliste bestimmt die zugehörige MDS-ID.
3. Die zentrale Patientenliste verschlüsselt die MDS-ID asymmetrisch mit dem öffentlichen Schlüssel (*MDS-DB-Pub*) der zentralen MDS-Datenbank.
4. Die verschlüsselte MDS-ID (*MDS-ID\**) wird an das lokale Datenmanagement zurückgegeben.
5. Verschlüsselte MDS-ID und MDAT werden vom lokalen Datenmanagement an die zentrale MDS-Datenbank übermittelt.
6. Die zentrale MDS-Datenbank entschlüsselt die MDS-ID mit ihrem privaten Schlüssel (*MDS-DB-Priv*).
7. Die zentrale MDS-Datenbank speichert MDS-ID und MDAT.
8. Die erfolgreiche Bearbeitung wird dem lokalen Datenmanagement zurückgemeldet.

Um Wörterbuchattacken zu vermeiden, wird durch Randomisierung im Verschlüsselungsalgorithmus sichergestellt, dass die wiederholte (d.h. zeitlich hintereinander erfolgende) Verschlüsselung der gleichen MDS-ID verschiedene Chiffre erzeugt.

Die übermittelten Daten können von den Administratoren von Brückenkopf und MDS-Datenbank im Rahmen von Wartungsarbeiten eingesehen werden. Ansonsten erfolgt durch den Upload selbst keine Sichtbarmachung, die Daten stehen mit dem Upload aber für die Zentrale Suche (siehe unten) zur Verfügung.

#### 4. Zentrale Suche

Mit der zentralen Suche können DKTK-Forscher den Datenbestand der zentralen MDS-Datenbank durchsuchen, um Standorte zu ermitteln, an denen Daten und Proben von DKTK-Patienten vorhanden sind, die für ein Forschungsvorhaben relevant sein könnten. Die zentrale MDS-Datenbank stellt dafür ein Webformular bereit, in dem Attribute der Meldedatensätze nach vorgegebenen Werten oder per Freitextsuche durchsucht werden können. Mehrere Suchattribute können durch logische Operatoren frei kombiniert werden.

Nach Ausführen der Suchanfrage enthält der anfragende Forscher folgende Informationen:

- Die Anzahl der Patienten bzw. Proben, die die Suchkriterien erfüllen.
- Die Altersverteilung der gefundenen Patienten in 10-Jahres-Intervallen.
- Die Geschlechtsverteilung (Anzahl Patienten männlichen / weiblichen / unbekanntes Geschlecht).
- Die Verteilung auf die Standorte.
- Kontaktdaten der Standorte, an denen die Daten erhoben wurden.

Die Datensätze selbst werden nicht übermittelt. Ein eventueller Zugriff auf Daten durch den anfragenden Forscher sowie die datenschutzrechtliche Grundlage dafür wird zwischen diesem und dem jeweiligen Dateneigentümer (=Standort) verhandelt und findet außerhalb der zentralen MDS-Datenbank statt.

#### 5. Dezentrale Suche

Wie die zentrale Suche dient die dezentrale Suche dem Auffinden von „passenden“ Patienten und Proben für ein Forschungsvorhaben, berücksichtigt im Gegensatz dazu aber auch Patienten ohne DKTK-Einwilligung. Da bei diesen angenommen werden muss, dass keine Rechtsgrundlage für die Übermittlung von Daten aus dem zuständigen Standort heraus gegeben ist, erfolgt hier keine automatische Übermittlung von Suchergebnissen.

Das Suchformular der dezentralen Suche wird vom Suchbroker für die dezentrale Suche bereitgestellt. Es erlaubt nicht nur die Suche nach den Meldedatensätzen, sondern allen im Metadata Repository (vgl. 4.1) abgelegten Begriffen; zusätzlich sind Ergänzungen im Freitext (Prosa) möglich. Der anzufragende Datensatz ist hier also prinzipi-

ell unbeschränkt. Die Suchanfrage wird zunächst gespeichert und dem anfragenden Forscher lediglich die Speicherung der Suchanfrage mitgeteilt. Die Teiler der Standorte rufen in regelmäßigen Abständen neu hinzugekommene Suchanfragen vom Suchbroker ab und ermitteln, welche Datensätze im lokalen Datenmanagement den Suchkriterien entsprechen. Der Inhalt der Anfrage sowie die gefundenen Datensätze können an jedem Standort von einer dazu berechtigten Person eingesehen werden. Diese kann nun den anfragenden Forscher kontaktieren, um eine mögliche Weitergabe von Daten oder Proben zu vereinbaren. Dieser Vorgang erfolgt wiederum außerhalb der CCP-IT und die damit verbundenen datenschutzrechtlichen Fragen müssen im Einzelfall von den beteiligten Personen geklärt werden.

## Abkürzungen

AnaDAT	s. xDAT
BMB	Biomaterialbank
CCP(-IT)	(IT der) Clinical Communication Platform
DKTK	Deutsches Konsortium für Translationale Krebsforschung
DKTK#ID	DKTK-weit verknüpfbarer Patientenidentifikator. # bezeichnet einen Standort, bspw. S <sub>MZ</sub> ID für Mainz. Eine DKTK#ID ist nur für den angegebenen Standort gültig, die verschiedenen DKTK#ID eines Patienten können aber in der zentralen Patientenliste einander zugeordnet werden.
ETL	Vorgang zur <u>E</u> xtraktion von Daten aus Quellsystemen, <u>T</u> ransformation in ein Zielformat, <u>L</u> aden in ein Zielsystem.
GTDS	Gießener Tumordokumentationssystem
ID	Primärer Patientenidentifikator (erster Stufe) im DKTK. Oberbegriff für DKTK#ID, MDS-ID und S#ID.
KN	Kontrollnummer
KNTKT	s. xTKT
MDAT	s. xDAT
MDR	Metadatenverzeichnis
MDS-ID	Patientenidentifikator in der MDS-Datenbank. Zeigt an, welche MDS-x-Datensätze der zentralen MDS zu einem Patienten gehören. Ist nur in der MDS-Datenbank gültig. Nur in der zentralen Patientenliste ist die Zuordnung zu den zugehörigen DKTK#IDs möglich.
MDS-x	Meldedatensatz für <ul style="list-style-type: none"> <li>- <u>K</u>linische Daten</li> <li>- <u>B</u>iomaterial</li> <li>- <u>S</u>tudien-Organisatorika</li> </ul>
OrgDAT	s. xDAT
PSN	Pseudonym höherer Stufe
S#ID	Primärer Patientenidentifikator (erster Stufe) im DKTK. # bezeichnet einen Standort, bspw. S <sub>MZ</sub> ID für Mainz. Nur lokal am bezeichneten Standort gültig, nicht übergreifend vergleichbar.
xDAT	Oberbegriff für Daten einer bestimmten Klasse: <ul style="list-style-type: none"> <li>- IDAT: direkt identifizierende Daten (Vorname, Nachname, Geburtsdatum, ...)</li> <li>- MDAT: klinische Daten (Obermenge des MDS-K)</li> <li>- AnaDAT: (Analyse-)Daten zur Bioprobe (Obermenge des MDS-B)</li> <li>- OrgDAT = (organisatorische) Verwaltungsdaten zur Bioprobe</li> </ul>
xTKT	(meist temporäres) Ticket für ein Datenelement