

miracum



**Unabhängige
Treuhandstelle**

UNIVERSITÄTSMEDIZIN GREIFSWALD

FÖDERIERTES RECORD LINKAGE: ANFORDERUNGEN UND PRINZIPIEN

09.09.2019

Christopher Hampf
Unabhängige Treuhandstelle
der Universitätsmedizin Greifswald



- i. Rechtliches und Organisatorisches zur THS Greifswald
- ii. Record Linkage mit E-PIX und Pseudonymverwaltung mit gPAS
- iii. Personen- und Identitätenkonzept
- iv. Föderiertes Record Linkage mit E-PIX
- v. Föderiertes Record Linkage der MI-I (MIRACUM)





Beschlussvorlage Seite 1 von 3

Drucksachen Nr:

033 / 2019

Universitätsmedizin Greifswald
Körperschaft des öffentlichen Rechts



Beschlussvorlage für den Vorstand

Sitzung des VS am:
03.04.2019

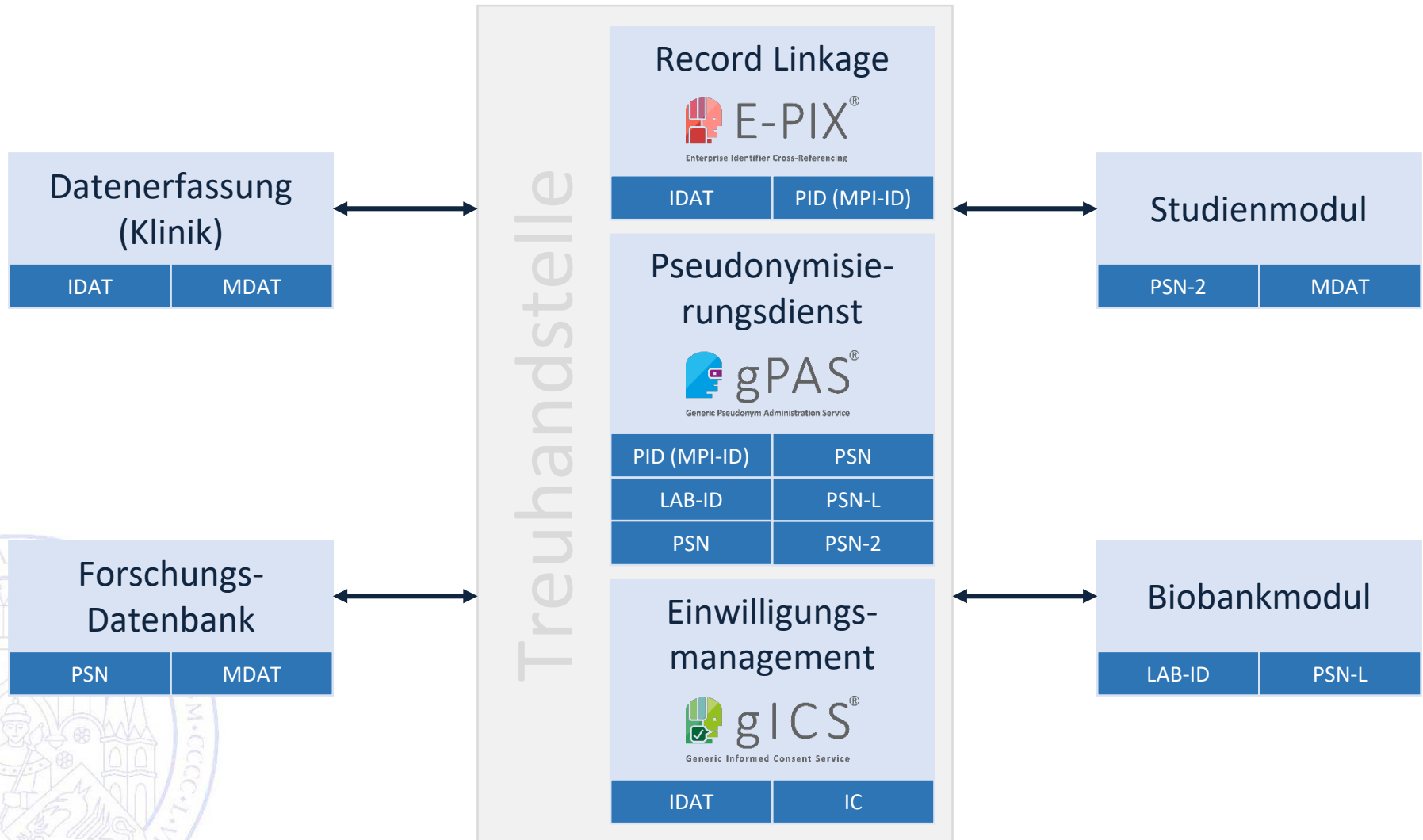
Betreff:

Weiterführung des Betriebs der unabhängigen Treuhandstelle.
Aktualisierung organisatorischer Aspekte der bereits per früherem Vorstandsbeschluss eingerichteten Treuhandstelle der Universitätsmedizin Greifswald

Die Universitätsmedizin Greifswald hat mit der ersten Beschlussvorlage vom **22.04.2014** eine zentrale unabhängige Einrichtung zum Betrieb einer Treuhandstelle errichtet. Die Treuhandstelle hat den Status einer Core Unit der Universitätsmedizin Greifswald. Organisatorisch wird sie formal dem Wissenschaftlichen Vorstand zugeordnet. Die Treuhandstelle bildet organisatorisch einen eigenständigen, räumlich von der Abteilung ICM-VC abgegrenzten Funktionsbereich. Mitarbeiter der Treuhandstelle werden i.d.R. durch Mandanten der Treuhandstelle finanziert. Mandanten sind Projekte oder Einrichtungen, die die Funktionen der Datentreuhänderschaft an die Treuhandstelle übertragen und die innerhalb der Treuhandstelle unabhängig voneinander betrieben und verwaltet werden.



Informationelle Gewaltenteilung (nach TMF-Leitfaden mit MOSAIC-Tools)





- i. Verwaltung von Personen und Identitäten
 - ii. Datenzusammenführung aus div. Quellen (z.B. SAP, M-KIS, SecuTrial, LIMS)
 - iii. Umgang mit fehlerhaften bzw. unvollständigen Daten
 - iv. Unterstützung bei der Auflösung von möglichen Synonymfehlern (Dublettenerkennung)
 - v. Konfigurierbare Matching-Parameter
 - vi. Protokollierung von Systemprozessen und wesentlichen Systementscheidungen
 - vii. Mandantenfähig
 - viii. Import und Export (CSV-Format)
- Performance durch Blocking und Caching

Possible matches

Person 1				Person 2			
Firstname	Lastname	♀	Birthdate	Firstname	Lastname	♀	Birthdate
Klaus	Schmidt	M	12.04.1970	Klaus-Dieter	Schmidt	M	21.04.1970
Max	Meier	M	01.01.1900	Max	Meier	M	01.01.1900

Occurred at: 06.09.2019 08:18:53, 29.03.2019 14:13:08

Compare data

Field	Person 1	Person 2
MPI	1001000000059	1001000000066
Last update	06.09.2019 08:18:27	06.09.2019 08:18:53
Local identifiers	0	0
External date		
Firstname	Klaus	Klaus-Dieter
Lastname	Schmidt	Schmidt
Title		
Middlename		
Birthname		
Gender	Male	Male
Birthdate	12.04.1970	21.04.1970
Birthplace		
Nationality		



E-PIX®

Enterprise Identifier Cross-Referencing

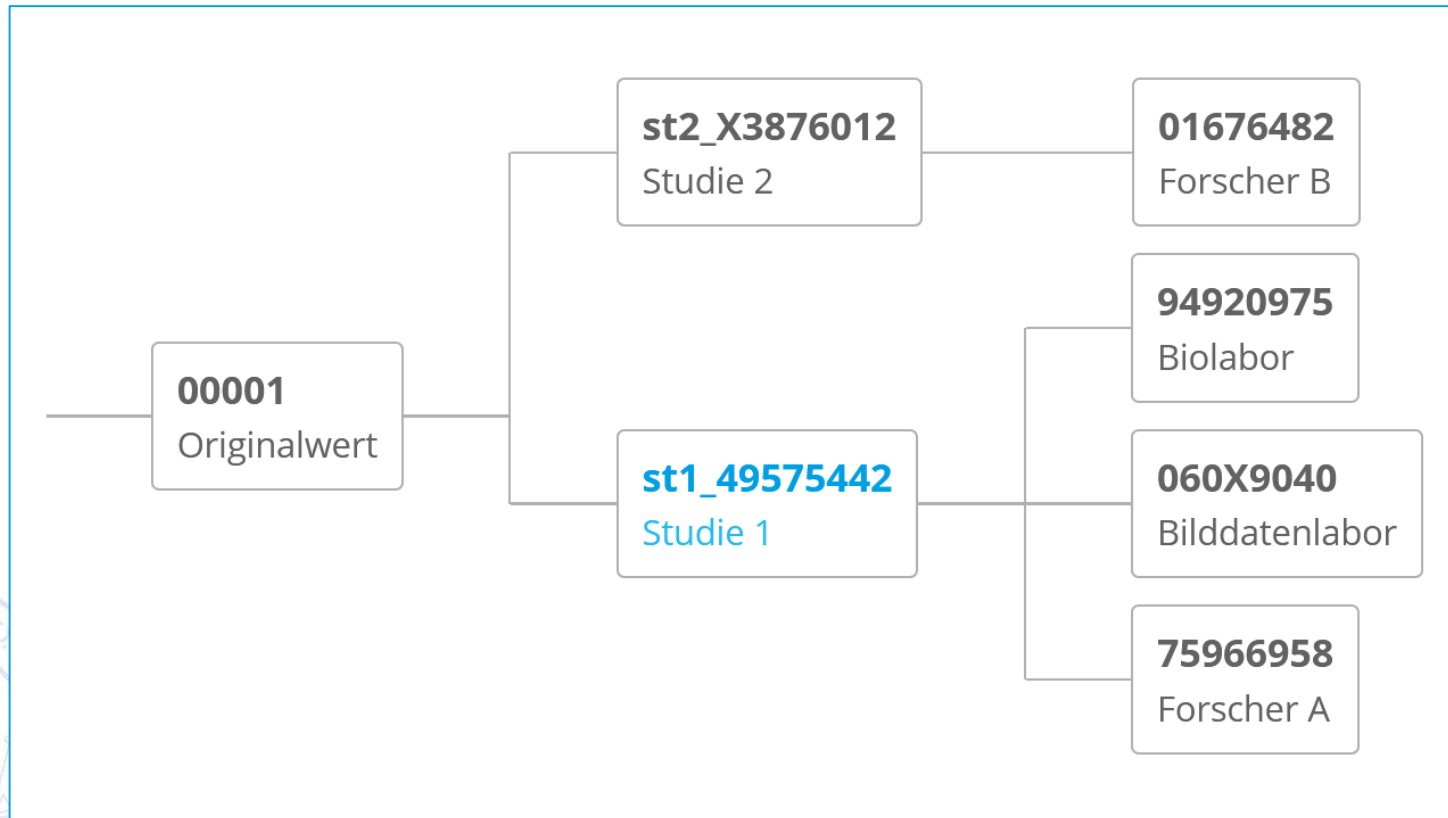


i. Funktionsumfang (Auszug)

- Anlegen neuer Pseudonyme durch die Benutzeroberfläche oder den Import von Eingabewerten
- Zuordnung von Pseudonymen zu beliebigen Eingabewerten
- Anonymisierung durch Löschung der zugeordneten Daten
- De-Pseudonymisierung durch Suchmechanismen / Filterfunktionen
- Löschung von Pseudonymen (dabei werden sowohl Pseudonym als auch zugeordneter Originalwert gelöscht)
- Erstellung und Anzeige von Pseudonym-Hierarchien
- Stapelverarbeitung
- Validierung von Pseudonymen
- Umfassende Konfigurierbarkeit

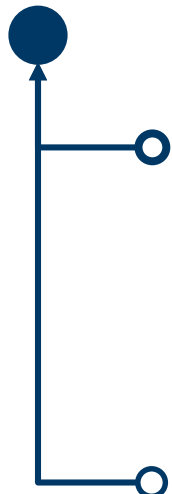


Generic Pseudonym Administration Service

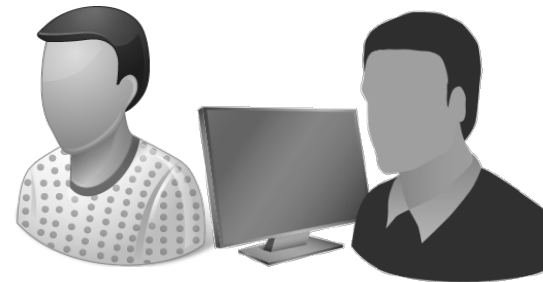




MPI-ID 1001000033333



Klaus-Dieter Schmidt
Pat-ID 00001234
Fall-Nr. 20170123
Fall-Nr. 20180543



Klaus Schmidt
LIMS-ID 12345





- i. Probabilistisches Verfahren für Matching von Datensätzen (Fellegi-Sunter)

$$R = P(M) / P(U) = \sum (p(m)) / \sum (p(u))$$

- ii. Verfahren für einzelne Felder: Levenshtein-Distanz, Phonetic, ...

- iii. Matchtypen

Perfect
Match

Match

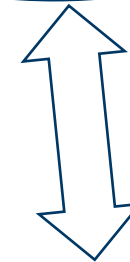
Possible
Match

No
Match





SITE A
Klaus Schmidt,
Geboren 21.04.1970
ID: 1001000011111





SITE B

Klaus Schmitt,
Geboren 12.04.1970

ID: 1001000011111



1001000011111



Klaus Schmitt



Klaus Schmitt

SITE C

Klaus-Dieter Schmidt,
Geboren 21.04.1970

ID: 1001000033333



1001000033333



Klaus-Dieter Schmidt



E-PIX® Possible matches

Persons
Possible Matches
Search / Edit
Add

Lists
Import
Export

Administration
Domains
Protocols
Statistic
Info

Current domain
Demo

Person 1			Person 2			Occured at
Firstname	Lastname	Birthdate	Firstname	Lastname	Birthdate	
Klaus	Schmidt	M 12.04.1970	Klaus-Dieter	Schmidt	M 21.04.1970	06.09.2019 08:18:53
Max	Maier	M 01.01.1900	Max	Meier	M 01.01.1990	29.03.2019 14:13:08

1-2 of 2
Create possible match

Compare data

Keep person 1		Keep person 2	
MPI	1001000000059	MPI	1001000000066
Last update	06.09.2019 08:18:27	Last update	06.09.2019 08:18:53
Local identifiers	0	Local identifiers	0
External date	Klaus	External date	Klaus-Dieter
Firstname	Schmidt	Firstname	Schmidt
Lastname		Lastname	
Title		Title	
Middlename		Middlename	
Birthname	Male	Birthname	Male
Gender	12.04.1970	Gender	21.04.1970
Birthdate		Birthdate	
Birthplace		Birthplace	
Nationality		Nationality	



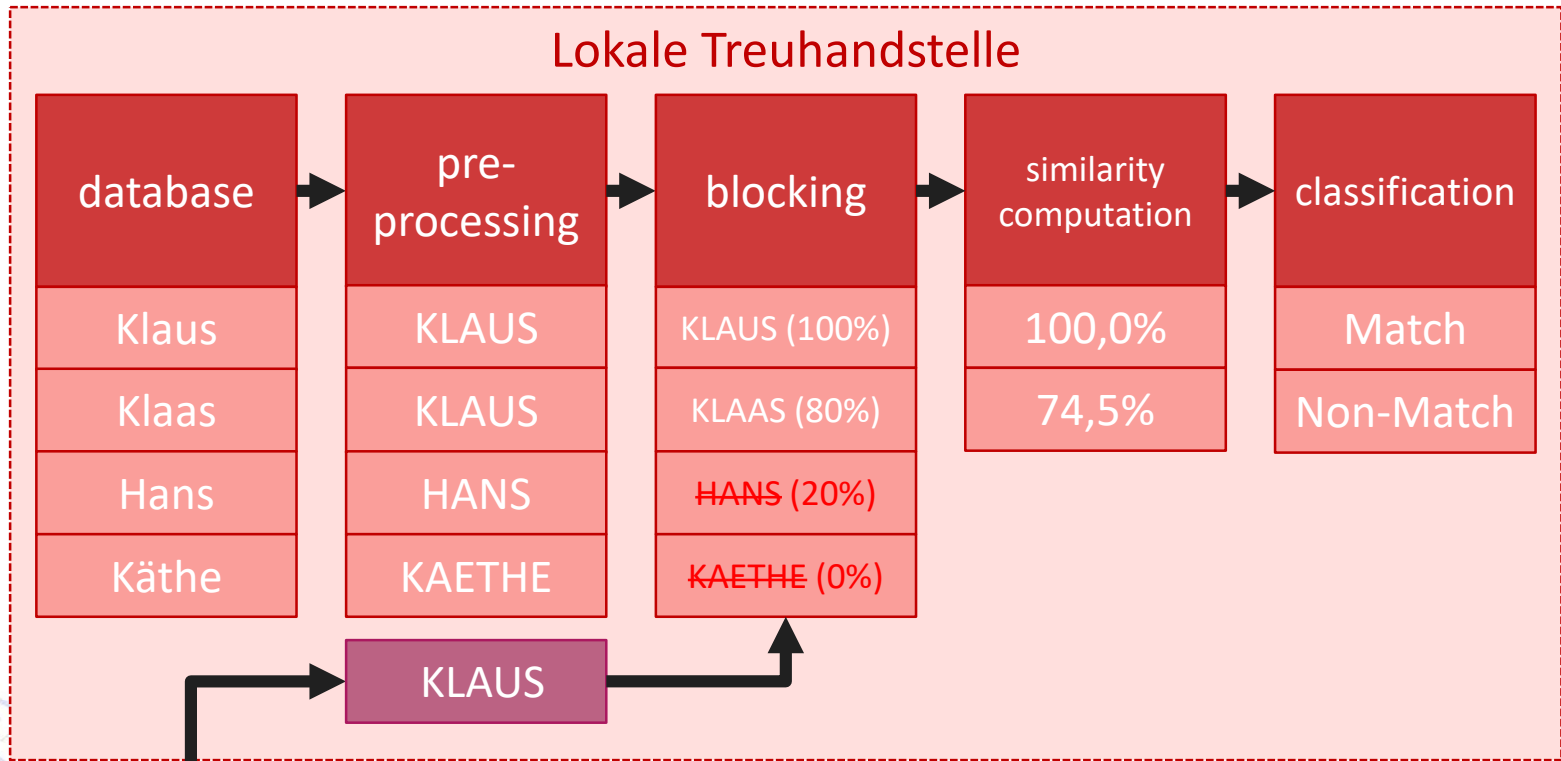
Klaus-Dieter Schmidt
1001000033333



Klaus-Dieter Schmidt

Klaus Schmidt
1001000011111

Klaus Schmitt



Patient erfasst



Klaus





Die Treuhandstelle in Zahlen

Stand Juli 2019



10.051.000

Pseudonyme individuell erzeugt



1.875.000

Teilnehmer sicher gespeichert



318.700

Einwilligungen digital erfasst



DZHK
DEUTSCHES ZENTRUM FÜR
HERZ-KREISLAUF-FORSCHUNG E.V.



miracum



RubiN

Regional ununterbrochen betreut im Netz

gefördertes Projekt
des Innovationsfonds
**Gemeinsamer
Bundesausschuss**



**KLINISCHES KREBSREGISTER
MECKLENBURG-VORPOMMERN**
Zentralstelle der Krebsregistrierung
Registerstellen gGmbH
Treuhandstelle



**MEDIZIN
INFORMATIK
INITIATIVE**



KAS+ Projekt

**Universitätsmedizin
GREIFSWALD**



DKMS
WIR BESIEGEN BLUTKREBS



CoBi
COLLABORATIVE
BIOBANK



DFPN
DEUTSCHES FORSCHUNGSPRAKTIKUM



NAKO
GESUNDHEITS-
STUDIE



EUROPÄISCHE UNION
Europäischer Fonds für
Regionale Entwicklung



MOSAIC



CAU
Christian-Albrechts-Universität zu Kiel



Interreg
Baltic Sea Region



**EUROPEAN
REGIONAL
DEVELOPMENT
FUND**
EUROPEAN UNION



CSB

Centrum für Schlaganfallforschung Berlin



PriVileG-M

Exzellenzinitiative M-V





























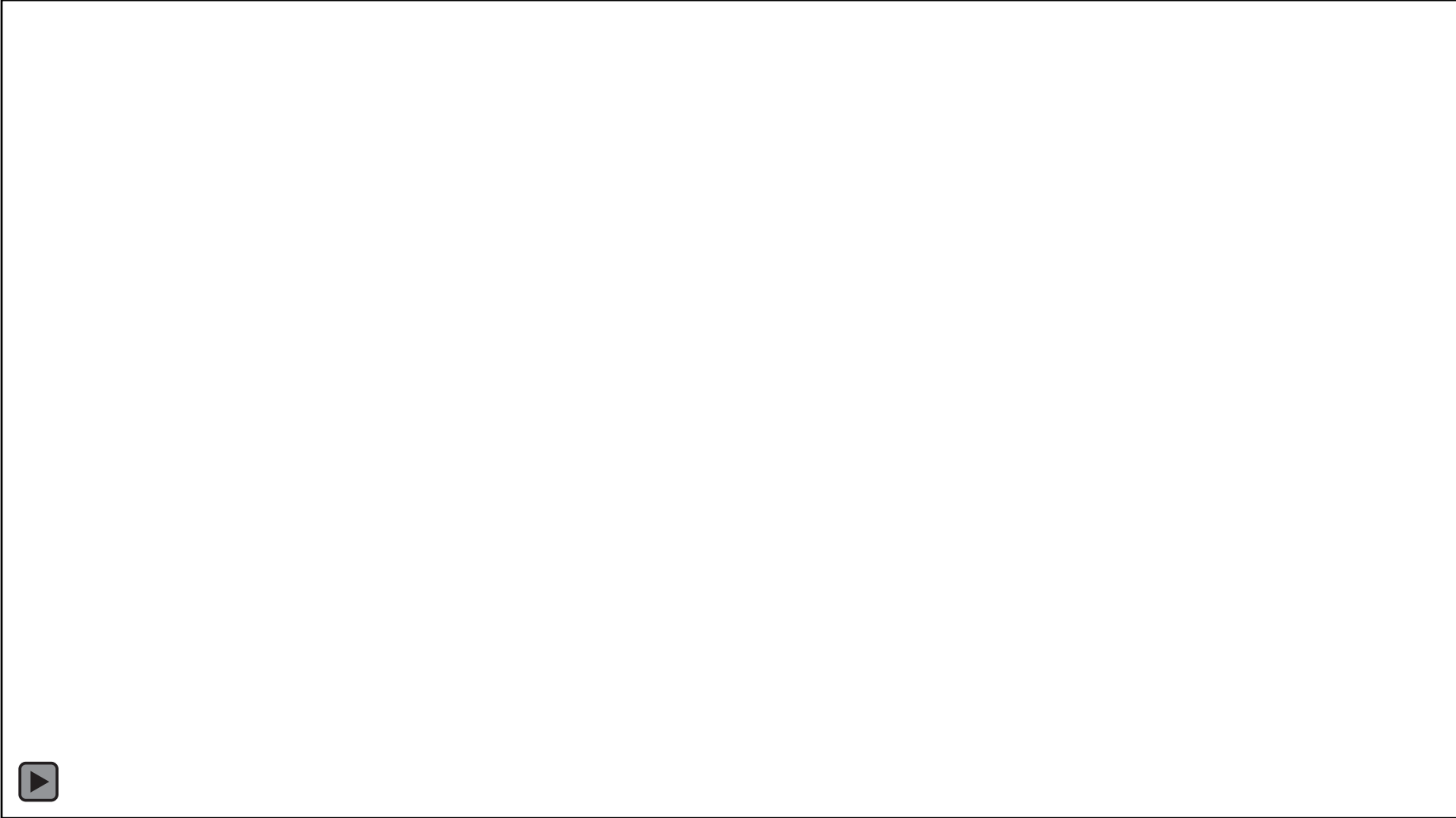
BFCC
Baltic Fracture Competence Centre

Übersicht eingesetzter Lösungen



Stand: 3. Sept. 2019

	Record Linkage	Pseudonymisierung	Einwilligungsmanagement
Dresden	 E-PIX®	 gPAS®	 gICS®
Erlangen	 E-PIX®	 gPAS®	 gICS®
Frankfurt	Mainzliste	Mainzliste	 gICS®
Freiburg	 E-PIX®	 gPAS®	 gICS®
Gießen	 E-PIX®	 gPAS®	 gICS®
Greifswald	 E-PIX®	 gPAS®	 gICS®
Magdeburg	 E-PIX®	 gPAS®	 gICS®
Mainz	Mainzliste	Mainzliste	 gICS®
Mannheim	 E-PIX®	 gPAS®	 gICS®
Marburg	 E-PIX®	 gPAS®	 gICS®



<https://www.youtube.com/watch?v=270VuBvzcj0>

Verteilung der durch die THS betreuten Studienzentren



DZHK
DEUTSCHES ZENTRUM FÜR
HERZ-KREISLAUF-FORSCHUNG E.V.



N = 140 Studienzentren
National: 120 Studienzentren
International: 20 Studienzentren

Internationale Studienzentren:

Barcelona – Hospital del Mar, Bialystok, Budapest – Military Hospital, Budapest – Teaching Hospital, Budapest – UK Semmelweis, Breslau (Wroclaw), Graz, Klagenfurt/Kopenhagen, Lissabon – Hospital de la Luz, Lissabon – Santa Maria Hospital, Lomza, Murska Sobota, Ózd, Pecs, Rom, Thessaloniki, Topolsica, Warschau, Valencia (de la Fe)



GESUNDHEITS-
STUDIE

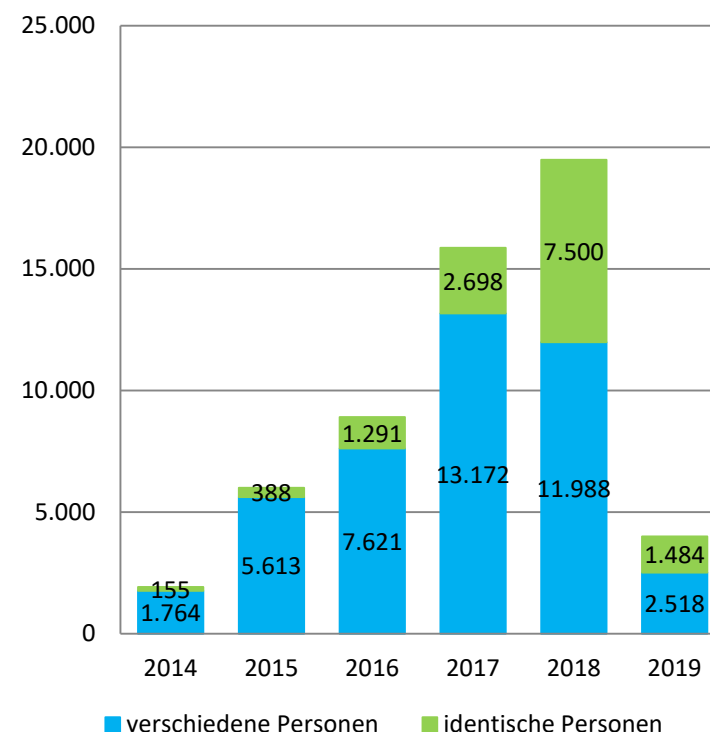
1.362.402 gespeicherte Personendatensätze:

- 2014: 61.328
- 2015: 255.552
- 2016: 322.249
- 2017: 349.097
- 2018: 312.711
- 2019: 61.465

In der Zeit entstandene mögliche Dubletten:

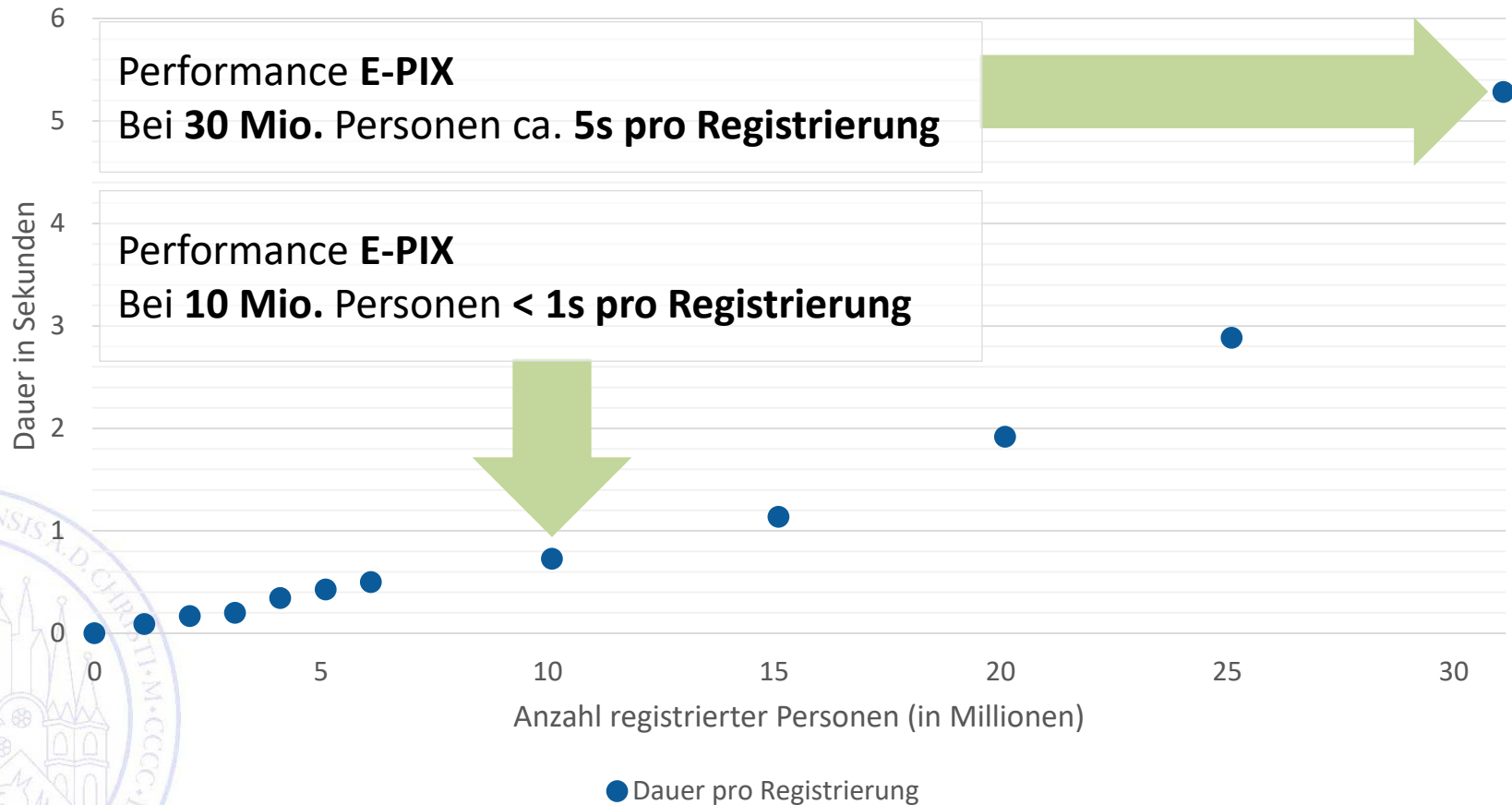
- 2014: 1.919
- 2015: 6.001
- 2016: 8.912
- 2017: 15.870
- 2018: 19.488
- 2019: 4.002

Entwicklung Dublettenprüfung







Durchschnittliche Dauer einer Personenregistrierung
im E-PIX bei unterschiedlich großen Datenbeständen












Possible matches

DE EN 





Persons

-  Possible Matches
-  Search / Edit
-  Add

Lists

-  Import
-  Export

Administration

-  Domains
-  Protocols
-  Statistic
-  Info


Current domain


Demo
▼

Possible matches: 2

Person 1				Person 2				Occured at
Firstname	Lastname	♀	Birthdate	Firstname	Lastname	♀	Birthdate	
Klaus	Schmidt	M	12.04.1970	Klaus-Dieter	Schmidt	M	21.04.1970	06.09.2019 08:18:53
Max	Maier	M	01.01.1900	Max	Meier	M	01.01.1990	29.03.2019 14:13:08

1-2 of 2 ⏪ ⏩ 1 ⏪ ⏩ per page:

 Create possible match

Compare data  Keep both persons

Keep person 1		Keep person 2	
MPI	1001000000059	MPI	1001000000066
Last update	06.09.2019 08:18:27	Last update	06.09.2019 08:18:53
Local identifiers	0	Local identifiers	0
External date		External date	
Firstname	Klaus	Firstname	Klaus-Dieter
Lastname	Schmidt	Lastname	Schmidt
Title		Title	
Middlename		Middlename	
Birthname		Birthname	
Gender	Male	Gender	Male
Birthdate	12.04.1970	Birthdate	21.04.1970
Birthplace		Birthplace	
Nationality		Nationality	

Quelle: Universitätsmedizin Greifswald, 2019

Alle Werkzeuge sind über eine responsive Web-Oberfläche bedienbar



Unabhängige
Treuhandstelle
UNIVERSITÄTSMEDIZIN GREIFSWALD



E-PIX[®]



gPAS[®]



gICS[®]

E-PIX[®] Add DEEN ?

Persons
Possible Matches
Search / Edit
Add

Fields marked with * are mandatory.
You can add an arbitrary number of contacts and local identifiers.

Identifying data

Title

Firstname *
Max

Lastname *
Meier

Middlename

Birthname
Schmitt

Birthdate (DD.MM.YYYY) *
18.05.1985

Birthplace
Hamburg

Gender *
Male

Prefix

Suffix

Ethnicity

Religion

Mother tongue

Nationality

Civil status

Administration

Domains

Protocols

Static

Info

Current domain
Demo

gPAS[®] Pseudonyms DEEN ?

Pseudonyms
Search / Create

Here you have the option to generate new pseudonyms or insert complete pseudonym pairs from external systems.
You can also search, copy, anonymise and delete a pseudonym after selecting a domain or open a pop tree.

Lists
Upload

Options
Domains
Static
About

New Pseudonym

Generate or Enter from external

Domains	Name	Original value	Pseudonym
Bildstellenlabor	sta_060509478	bw_700201446	
Bildstellenlabor	sta_242500239	bw_806000985	
Forscher 1	sta_526005308	bw_814877740	
Forscher 2	sta_901431933	bw_864713056	
Forscher 3	sta_941616259	bw_031441292	
Studie A			
Studie B			

Institute for Community Medicine, Greifswald - gPAS 1.8.1 - Demo

gICS[®] Consents DEEN ?

Documents
Consents
Withdrawals
Referrals
Search

Here you can add new consents or display existing ones.
By clicking right on a row, you will open additional options.

New Consent

Fill in or Upload and parse

Forms
Templates
Modules
Policies

Options
Domains
Import / Export
Static
About

Active domain
Demo Deutschland

2 consents existent	Date	Signed IDs	Template	Version	Details	Scan
	2019-01-15 03:11:16 PM	S.40002	Erwältigung Demostudie	1.0		
	2019-01-15 03:07:11 PM	S.60001	Erwältigung Demostudie	1.0		

12 of 2

Website for Community Medicine, Greifswald - gICS 2.9.1 - Demo



Demosysteme

E-PIX: <https://www.ths-greifswald.de/e-pix/demo>

gPAS: <https://www.ths-greifswald.de/gpas/demo>

gICS: <https://www.ths-greifswald.de/gics/demo>

Standortübergreifendes förderiertes Record Linkage in MIRACUM

MI-I | NSG | Stand 31.03.2017

Roadmap – Meilensteine

1. Standards für die Patienteneinwilligung
 - M 1.1: Versionen des Mustertextes liegen consentiert vor

2. Vertrauens- bzw. Treuhandstellen

- M 2.1: Spezifikation der konsortieninternen Treuhandstellen vorhanden
- M 2.2: Anforderungspapier zur Föderierung von Treuhandstellen liegt vor

3. Regelwerk Use & Access

- M 3.1: Eckpunktepapier AG Data Sharing liegt vor
- M 3.2: Technische Unterstützung für Use&Access ist vorhanden

4. Methoden Datensicherheit und -schutz

- M 4.1: Initiale Datenschutzkonzepte liegen vor
- M 4.2: Nutzer sind identifizier- und authentifizierbar gemäß definierter Berechtigungsstufen
- M 4.3: Weiterführende Datenschutzkonzepte liegen vor

5. Semantische Interoperabilität

- M 5.1: Inhalte, Terminologien (Kerndatensatz abgestimmt, Kerndatensatz bereitgestellt, Erweiterungsmodule abgestimmt)
- M 5.2: Strukturen, Informationsmodell, Syntax (für Basismodul gestaltet, für Erweiterungsmodule gestaltet)
- M 5.3: Konzept für Metadaten liegt vor (Qualität/Verfügbarkeit/Provenance)
- M 5.4: Übergreifende nutzbare Dienste sind verfügbar (Terminologien, Metadaten)

6. Methoden des Datenaustauschs

- M 6.1: Übergreifende Abfragen u. Datenaustausch sind möglich (Basismodul, Erweiterungsmodule)
- M 6.2: Laufzeitumgebungen sind interoperabel gestaltet

7. Übergreifende Anwendungsfälle zur Demonstration des Nutzens

- M 7.1: Spezifikationen gemeinsamer Use Cases liegen vor
- M 7.2: Projectathons finden statt

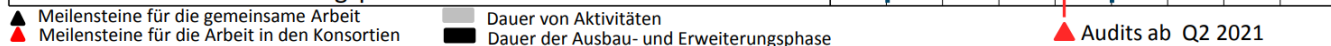
8. Einbeziehung von Patienten

- M 8: Workshop „Abstimmung des Vorgehens zur Einbeziehung von Patienten“

9. Stärkung von Forschung, Lehre und Weiterbildung

- M 9: Aktivitäten zur Stärkung von Forschung und Lehre (Workshop, Status Symposium)

10. Ausbau- und Erweiterungsphase



- Föderiertes Record Linkage (fRL) funktioniert grundsätzlich gleich zwischen Standorten eines Konsortiums und über Konsortiumsgrenzen hinweg (Jeder MII-Standort errichtet ein Datenintegrationszentrum (DIZ))
- Ziel ist die Datenübergabe, -nutzung und -auswertung mithilfe projektspezifischer, standortübergreifender und MI-I-weit eindeutiger Pseudonyme
- Arbeiten der TF DS zielen auf die Weiterentwicklung bereits bestehender Record Linkage Lösungen ab

- Die unterschiedlichen Ausprägungen einer mehrfach erfassten Person, werden dieser als Personenidentitäten im Ziel-Bestand zugeordnet
- Privacy Management Plug, welcher an jedem DIZ eingerichtet wird und eine harmonisierte Datenübermittlung zur federated Trusted Third Party (fTTP) ermöglicht
- Die fTTP unterstützt förderiertes Record Linkage bzw. SMPC/PPRL mittels Vergleich codierter IDAT und ermittelt damit Wahrscheinlichkeiten zur Zugehörigkeit zu einer Person
- Zusätzlich: In ausgewählten Anwendungsfällen (z.B. Seltene Erkrankungen), kann die Eindeutigkeit erforderlich sein, wodurch ein Abgleich mittels Klartext-IDAT erforderlich wird

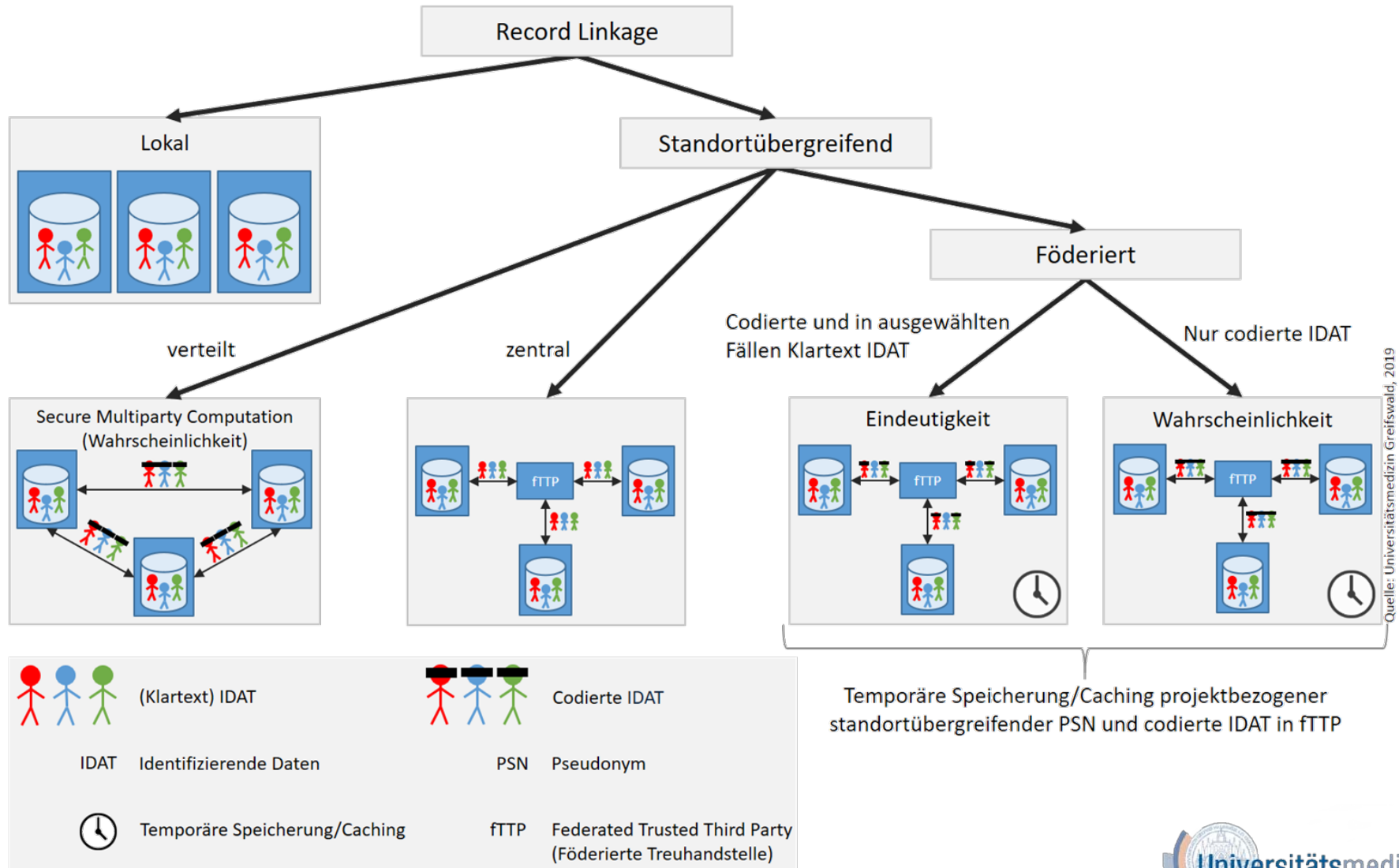
- Umgang mit fehlerhaften und unvollständigen Daten (Tippfehler, Zahlendreher, ...)
- Verwaltung von mehreren Identitäten, die zu einer Person zugehörig sind (unterschiedliche Ausprägungen von IDAT)
- Adäquate Performance für das Record Linkage in multi-millionen Datensätzen
- Pseudonymhierarchien

- Machbarkeitsuntersuchungen (Feasibility)
- Verteiltes Rechnen
- Kohorten-Identifikationen (z.B. für Seltene Erkrankungen)
- „Echte“ Datenzusammenführungen/Auswertungen
- Datenübergaben/-transfers, Data Sharing, Use & Access
- Sekundärdaten-Integration/-Nutzung
- ErgebnISRückführungen, Integration von Ergebnisdaten
- Re-Kontaktierungen
- Langzeitliche epidemiologische Analysen, Nachverfolgungen

Im Rahmen von föderiertem Record Linkage gilt:

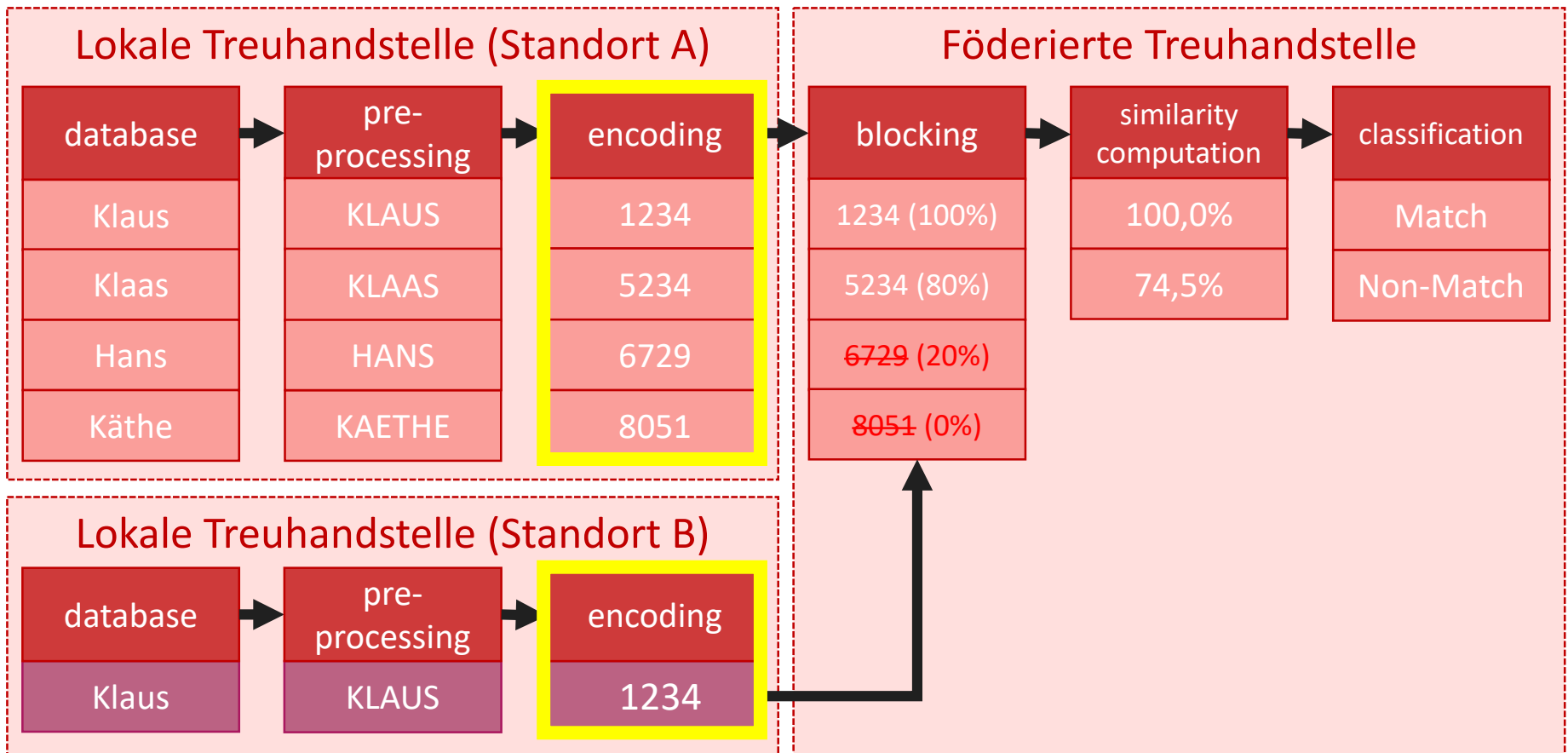
- Die **Wahrscheinlichkeit** gibt die Übereinstimmung zweier codierter IDAT-Datensätze an. Je ähnlicher sich diese sind, desto wahrscheinlicher gehören beide Datensätze zu einer Person.
- Die **Eindeutigkeit** gibt an, ob ein IDAT-Datensatz eindeutig einer Person zugeordnet werden kann.
- Privacy Preserving Record Linkage (PPRL) kann nur **Kandidaten** finden. Ein Kandidat besteht aus zwei codierten IDAT-Datensätzen, die aufgrund ihrer Ähnlichkeit wahrscheinlich zu einer Person gehören.
- Die hinreichende Wahrscheinlichkeit der Zugehörigkeit zweier codierter IDAT-Datensätzen zu einer Person hängt z.B. von gesetzten Schwellwerten ab

Systematisierung Record Linkage



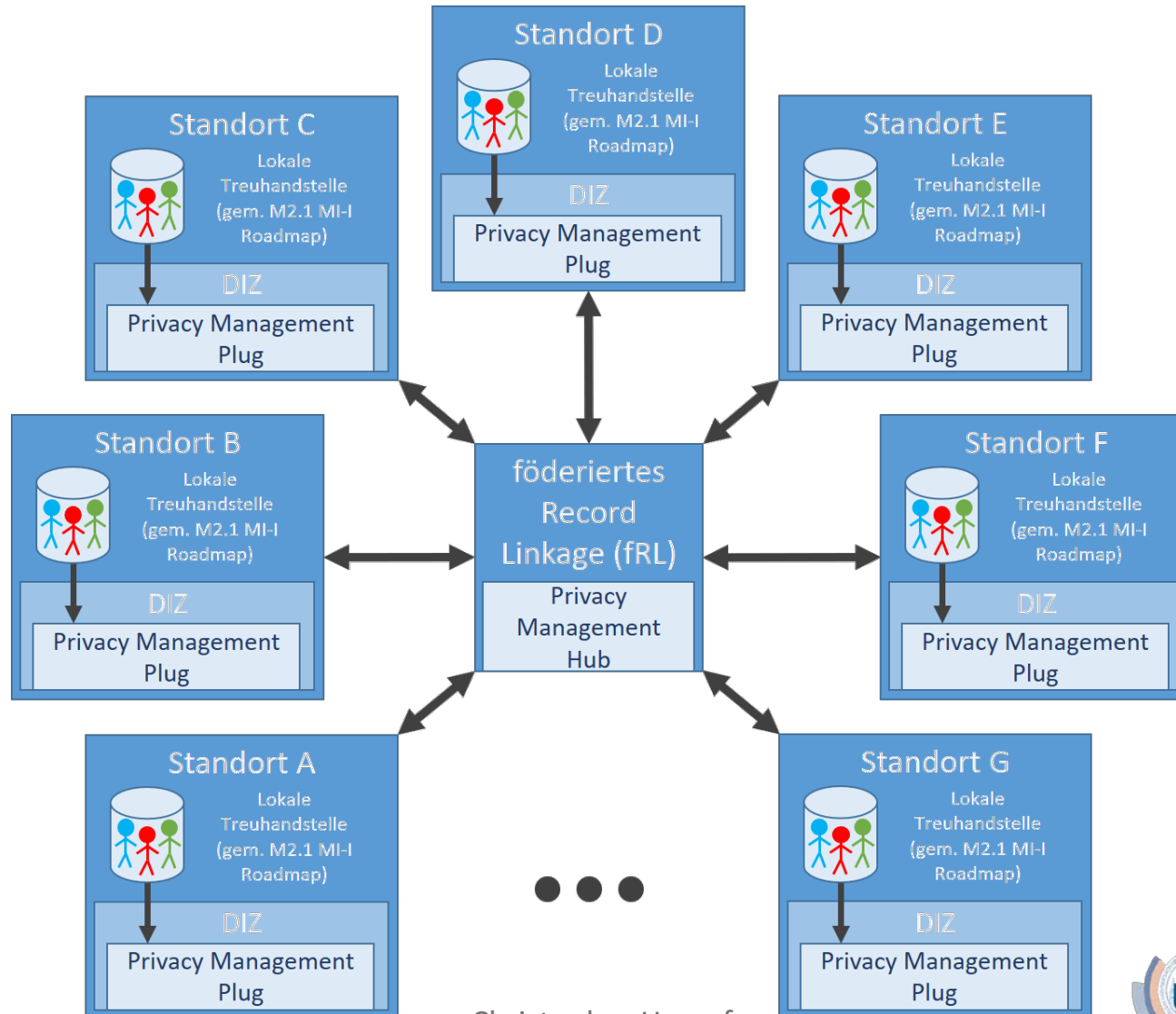
Quelle: Universitätsmedizin Greifswald, 2019

Föderiertes Record Linkage mit Vergleich von codierten IDAT

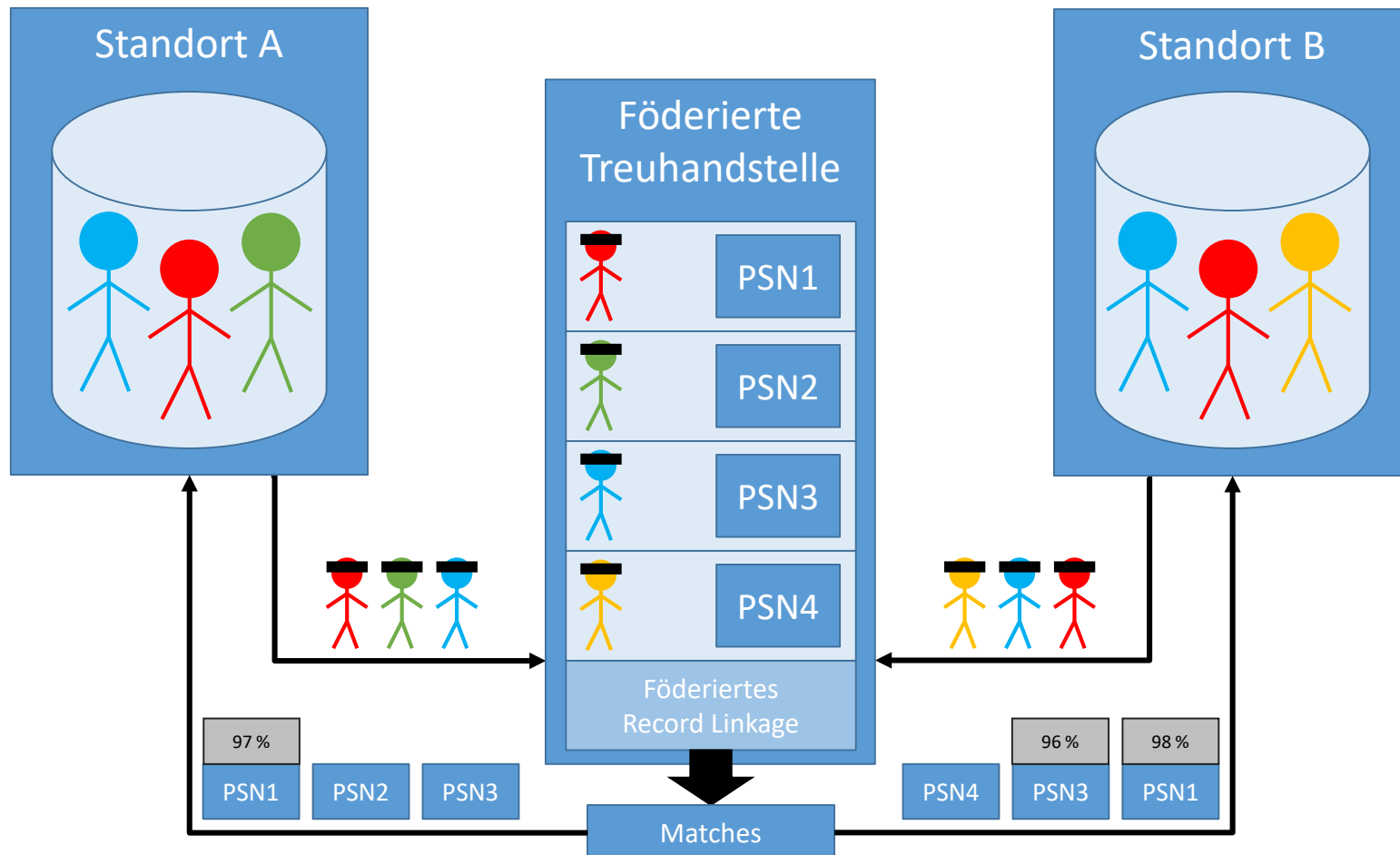


Quelle: Universitätsmedizin Greifswald, 2019

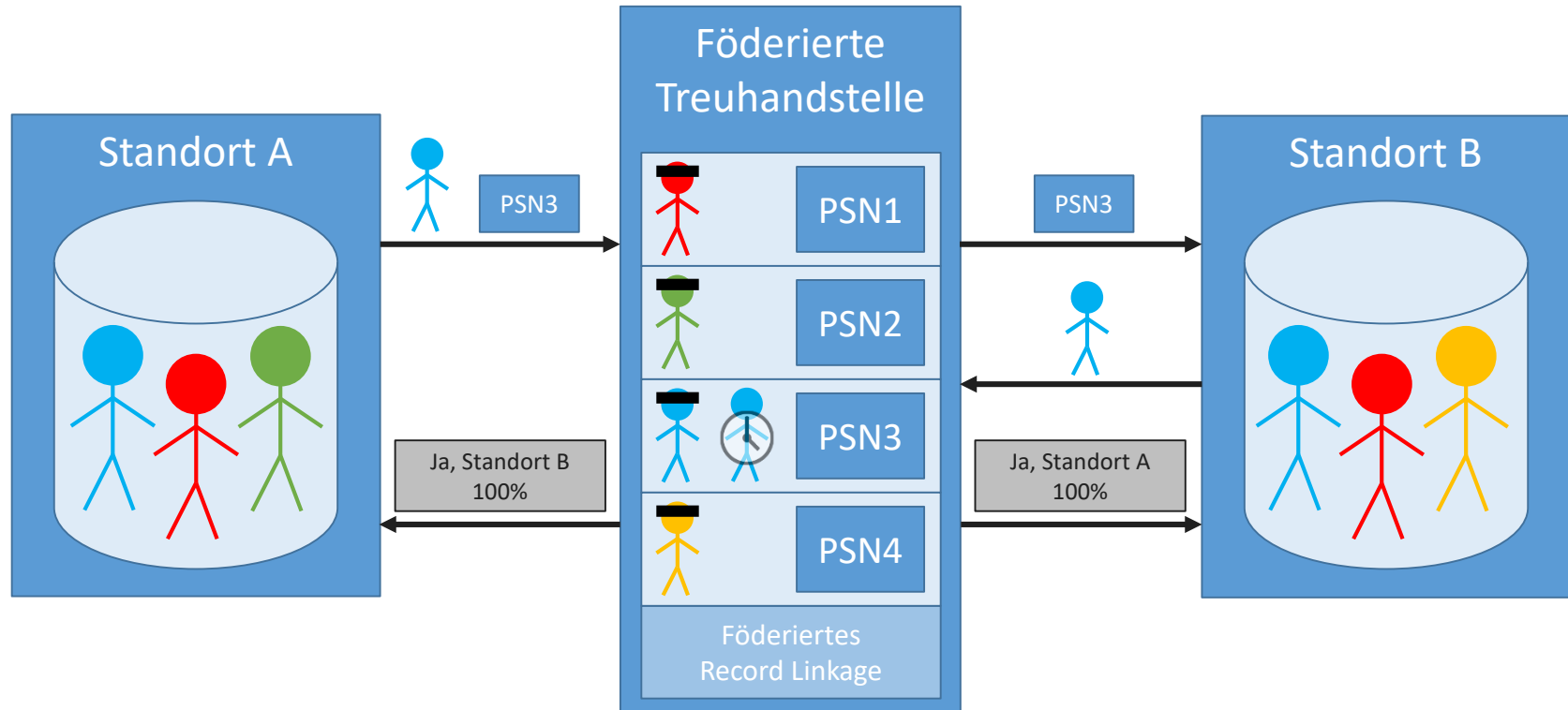
Konzept: Privacy Management Hub



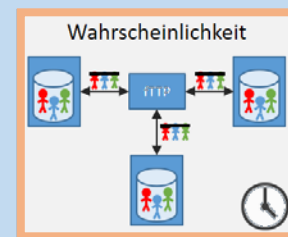
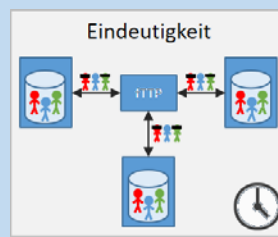
Föderiertes Record Linkage basierend auf Wahrscheinlichkeiten und PSN-Vergabe



Diskussion: Föderiertes Record Linkage basierend auf Eindeutigkeit



Konzept fRL (Wahrscheinlichkeiten)



Standort A

Lokale Treuhandstelle



Lokales Record Linkage

DIZ

Privacy Management Plug



Interface

fTTP

IDAT	PSN	Transfer
IDAT	PSN	Transfer
	PSN1	TPSN10
	PSN2	TPSN11

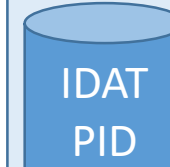
Föderiertes Record Linkage mit
cod. IDAT
(Wahrscheinlichkeit)

Privacy Management Hub

Standort B

Lokale Treuhandstelle

Mainzelliste



Lokales Record Linkage

DIZ

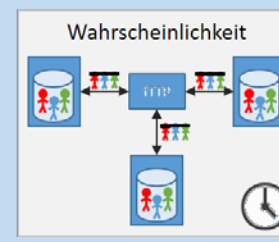
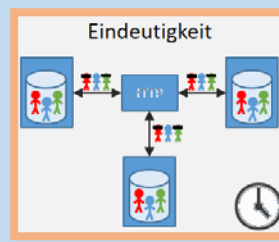
Privacy Management Plug



Interface

IDAT → cod. IDAT
(z.B. mit Bloom-Filter)

Konzept fRL (Eindeutigkeit)



Standort A

Lokale Treuhandstelle



Lokales Record Linkage

DIZ

Privacy Management Plug



Interface

fTTP

IDAT	PSN	Transfer
IDAT	PSN	Transfer
	PSN1	TPSN10
	PSN2	TPSN11

Föderiertes Record Linkage mit IDAT (Eindeutigkeit)

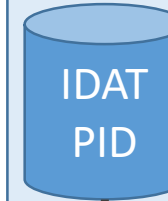


Föderiertes Record Linkage mit cod. IDAT (Wahrscheinlichkeit)

Privacy Management Hub

Standort B

Lokale Treuhandstelle Mainzelliste



Lokales Record Linkage

DIZ

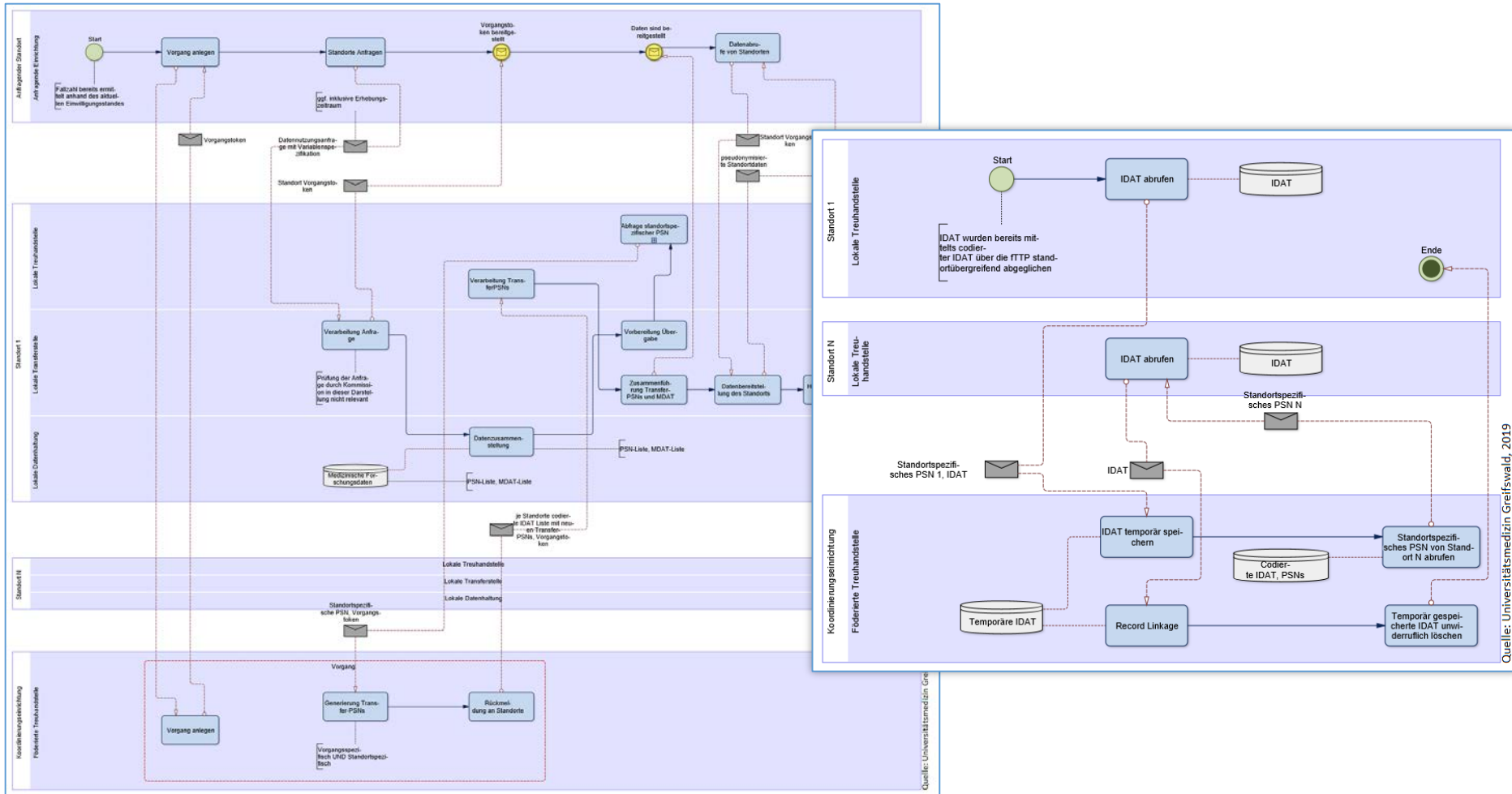
Privacy Management Plug



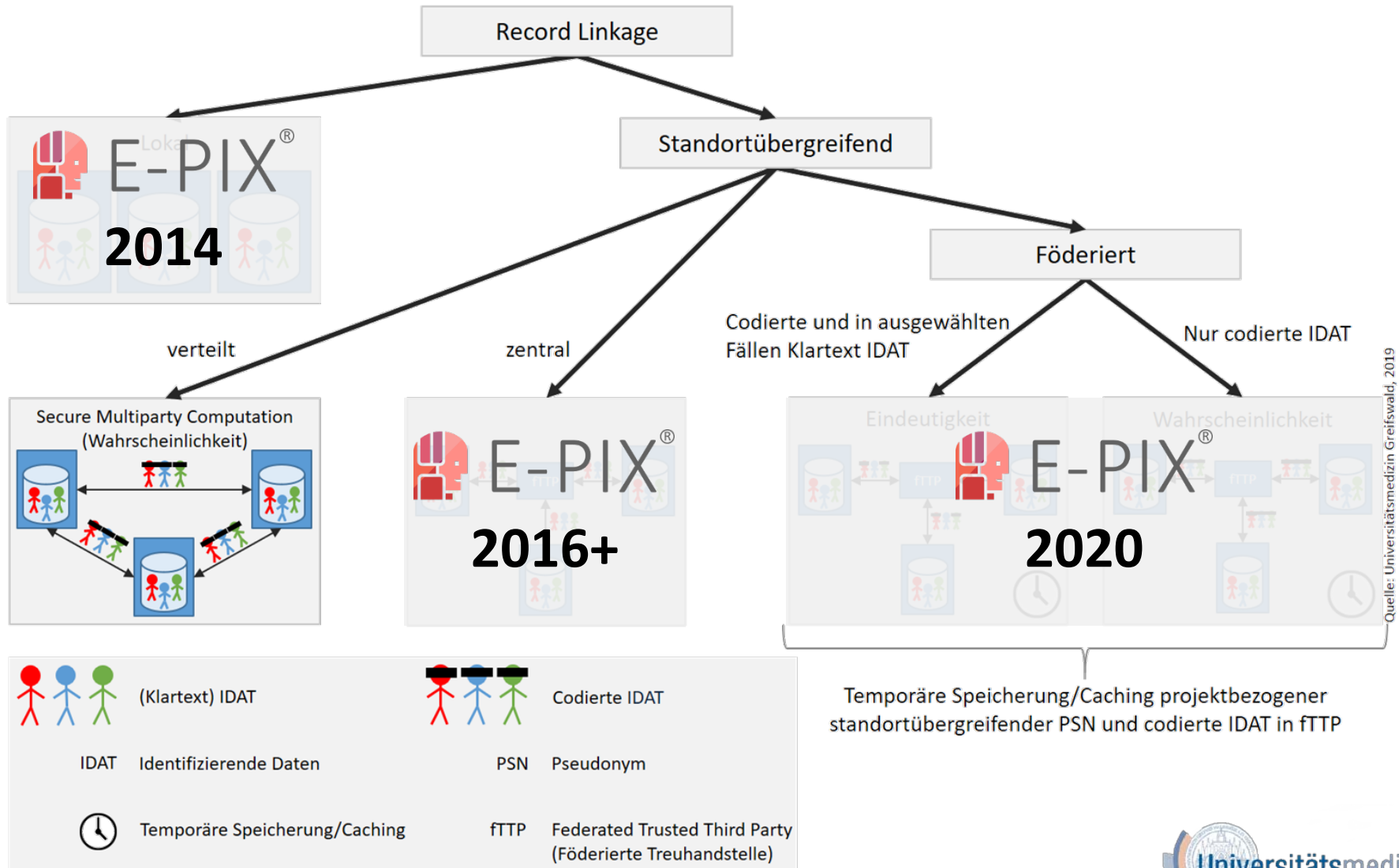
Interface

IDAT → cod. IDAT
(z.B. mit Bloom-Filter)




Vorschlag zu Prozessen der fTTP



Systematisierung Record Linkage



Quelle: Universitätsmedizin Greifswald, 2019


	(Klartext) IDAT		Codierte IDAT
IDAT	Identifizierende Daten	PSN	Pseudonym
	Temporäre Speicherung/Caching	fTTP	Federated Trusted Third Party (Föderierte Treuhandstelle)

- Erarbeitung von Prozessen und Schnittstellen zum standortübergreifendem Record Linkage durch die Taskforce Datenschutz
- Aufbauend auf bewährten Methoden wird der E-PIX bezüglich föderiertem Record Linkage weiterentwickelt; ebenso weitere verfügbare Lösungen (Harmonisierung)
- Konsortialübergreifende Pilotierung und Evaluierung von Lösungen
- M2.2: Anforderungspapier zur Föderierung von Treuhandstellen

Fragen?

Kontakt

Christopher Hampf
Universitätsmedizin Greifswald
Institute für Community Medicine
Ellernholzstr. 1-2
17475 Greifswald

 +49 (3834) 86-7851

 christopher.hampf@uni-greifswald.de

 ths-greifswald.de

KAS+ System Architecture at University Medicine Greifswald

