

Der PID-Generator der TMF

TMF-Workshop „Tools zum ID-Management in der klinischen Forschung“

Berlin, 24. September 2010

Prof. Dr. Klaus Pommerening, Dr. Murat Sariyar

Universitätsmedizin Mainz, IMBEI

KN Pädiatrische Onkologie und Hämatologie

TMF-AG Datenschutz



- ↪ Erzeugung nichtsprechender Identifikatoren, (Pseudonyme 1. Stufe), geeignet als
 - ↪ PID (= PID_V) in Versorgungsdatenbank,
 - ↪ SIC in Studiendatenbank (oder PID_S),
 - ↪ Input des Pseudonymisierungsdienstes für Forschungsdatenbank.
- ↪ Auffangen von Fehlern in IDAT (fehlertolerantes Matchen).
- ↪ Einbindung in verschiedene medizinische Forschungsszenarien als eine zentrale Komponente des ID-Managements.

- ↪ Web-Formular für PID-Anforderung durch Behandler/
Studienärzte;
 - ↪ Übertragung mit „Copy & Paste“.
- ↪ Batch-Verarbeitung auch möglich
 - ↪ durch Server-Administrator,
- ↪ Einbindung in vorhandene Erfassungsprogramme:
 - ↪ SOAP-Schnittstelle als Ansatzpunkt.
- ↪ Weitreichende Konfigurationsmöglichkeiten
 - ↪ Datenbankschema, Datenbankbindung, Templates für
Benutzungsoberfläche, Match-Verfahren, Meldungstexte,
Protokollierung, ...



Anforderung eines Patienten-Identifikators (GPOH-PID)

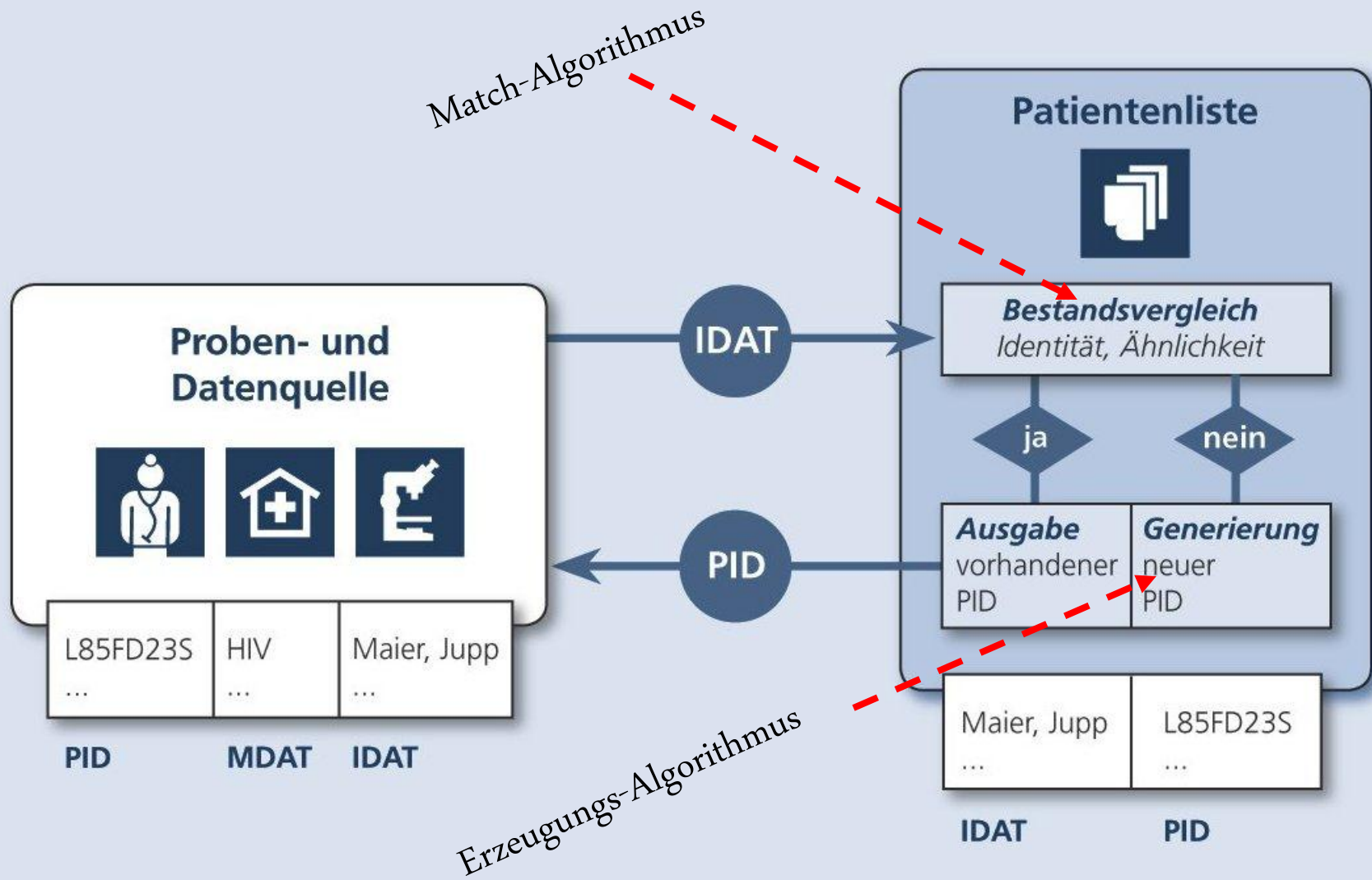


[Erklärung/Hilfe](#) [Vor der ersten Verwendung unbedingt lesen!]

Identifizierende Angaben		Wie sicher ist der Name?	
Nachname:	<input type="text"/>	<input checked="" type="radio"/> sicher	<input type="radio"/> unsicher
früherer Nachname:	<input type="text"/>	Vorname:	<input type="text"/>
		Geburtsdatum	TT: <input type="text"/> MM: <input type="text"/> JJJJ: <input type="text"/>
Ergänzende Angaben			
Geschlecht:	<input type="radio"/> weiblich <input type="radio"/> männlich <input checked="" type="radio"/> unbekannt		
Postleitzahl:	<input type="text"/>	Wohnort:	<input type="text"/>
		Staat:	<input type="text"/>

Bevor Sie das Formular abschicken, vergewissern Sie sich bitte noch einmal, ob alle Einträge korrekt sind.

Falls Sie als Reaktion nicht einen PID oder eine verständliche Fehlermeldung zurück erhalten, wenden Sie sich per [E-Mail](mailto:PIDservice@gpoh.de) an PIDservice@gpoh.de.



*Richtige Zuordnung bei fehlerhaften IDAT (fast) nur **vor** Pseudonymisierung möglich.*

↳ Daher integraler Bestandteil des pseudonymen ID-Mgt.

Logisches Matchen

↳ Erkennen von (z. B.) Namensänderung, Namenszusatz

↳ Wenn möglich, KV-Nummer verwenden.

Matchen mit »*unsicherem*« Namen:

↳ Zusatzdaten und phonetische Codes werden mitverwendet,

↳ evtl. Warnhinweis.

Homonym- vs. Synonymfehler

↳ Stochastische/ KI-Matchverfahren getestet,

↳ „klassisches“ Record Linkage leicht überlegen.

↳ Fehler nie ganz auszuschließen,

↳ Phonetik reduziert Synonymfehler.

PIDs maschinenlesbar (AES-Chiffre von lfd. Nummer)
oder menschenlesbar (8-Zeichen, Faldum-Code).

Unauflösbare Fehler erfordern Administrator-Eingriff

↳ Direktzugriff auf Datenbank.

Netzspezifische PIDs durch frei gewählten geheimen Schlüssel.

↳ Dieser darf nie mehr geändert werden.

Match-Verfahren prinzipiell jederzeit austauschbar, optimierbar.

Lauffähig unter Unix/Windows mit PostgreSQL oder ODBC.

- ↪ Software am IMBEI entwickelt – Entwickler und Betreuer:
 - ↪ Markus Wagner (bis 2003)
 - ↪ Jutta Glock (Moormann) (2003-2005)
 - ↪ Murat Sariyar (seit 2005)
- ↪ Für verschiedene Kompetenznetze zentraler Web-Service am IMBEI,
 - ↪ auf dediziertem Server.
- ↪ PID-Generator in Kompetenznetz POH seit 2002 (> 60 000 PIDs).
- ↪ Die 44 248 „Altfälle“ des Deutschen Kinderkrebsregisters im Batch-Lauf eingespeist.

Installationsaufwand für Web-Schnittstelle

- ↪ Apache, SSL (Zertifikate, PKI)
- ↪ Berechtigungen über LDAP, Zertifikate oder .htaccess
- ↪ CGI-Schnittstelle (HTML-Templates)
- ↪ Konfigurationsdateien des PID-Generators anpassen
- ↪ Datenbank
- ↪ Testen

=> Ein paar Tage

Installationsaufwand für Webservice-Anbindung

- ↪ Web-Schnittstelle
- ↪ Javaumgebung
- ↪ Tomcat, Axis, SOAP
- ↪ WebObjects für Unix, HTTP-Adaptor
- ↪ PIDDispatcher
- ↪ Testen

=> Kann auch mal länger dauern

GPOH-PID-Anfragen: ca. 320 pro Monat,
 Neue PIDs: ca. 230 pro Monat.

Häufigkeit manueller Eingriffe: im Schnitt 4x pro Monat

Vorgehen bei Duplikaten

- ↪ sub-Feld des „falschen“ Falles mit richtigem PID belegen
- ↪ Fall ist damit unsichtbar

Kein PID erzeugbar?

- ↪ Wie oben (eigene PID in sub) oder Brachialmethode: mittels psx enc Verschlüsselung erzeugen und Feld in DB überschreiben
- ↪ CAVE: Namensteile und Phonetik über Test-DB berücksichtigen

Erklärungsbedarf: im Schnitt 2x pro Monat