



Ansätze zu konsistenter Pseudonymfindung für klinische Dokumente

Stefan Schulz, Averbis GmbH

9.4.2021

WAS HEISST KONSISTENTE PSEUDONYMISIERUNG?

- Zum Begriff "Pseudonymisierung"
Ersatz von Identifikatoren
Hier Ersatz aller PHI
- Sicherheit: keine Re-identifizierung möglich
- Plausibilität: das Ergebnis soll von einem Originaldokument in der Struktur nicht unterscheidbar sein
- Konsistenz: Durch Pseudonymisierung dürfen innerhalb von Dokumenten eines Patienten keine Zusammenhänge (Personen, Zeitangaben) verloren gehen

Disclaimer: die hier geäußerten Überlegungen sind Diskussionsbeiträge. Sie sind nicht in Averbis-Produkten implementiert

HIPPAA PHI (Protected Health Information)

Names (Full or last name and initial)
All geographical identifiers smaller than a state (...)
Phone Numbers
Fax numbers
Email addresses
Social Security numbers
Medical record numbers
Health insurance beneficiary numbers
Account numbers
Certificate/license numbers
Vehicle identifiers
Device identifiers and serial numbers;
Web Uniform Resource Locators (URLs)
Internet Protocol (IP) address numbers
Biometric identifiers, including finger, retinal and voice prints
Full face photographic images and any comparable images
Any other unique identifying number, characteristic, or code (...)

Original

Krankenhaus der Samariter Holzhausen

Röntgenabteilung, Vorstand Prim. Univ. Prof. Dr.Dr. Gotthard Vogler

CT Abdomen und kl. Becken

Name: Mustafa Üstün, * 21.06.67

Aufnahmezahl: 1933309807

Abteilung: Chirurgie

Station: A31. OG. Viszeralchirurgie B /

Zi: 119

dikt. Arzt: OA Dr. Huber Karina

WinA. 06/07/2011

Getaggt

<location>Krankenhaus der Samariter</location>

<location>Holzhausen</location>

Röntgenabteilung, Vorstand <name>Prim. Univ. Prof. Dr.Dr. Gotthard Vogler</name>

CT Abdomen und kl. Becken

Name: <name>Mustafa Üstün</name>, *
<date>21.06.67</date>

Aufnahmezahl: <id>1933309807</id>

Abteilung: Chirurgie

Station: <division>A31. OG. Viszeralchirurgie B</division> /

Zi: 119

dikt. Arzt: <name>OA Dr. Huber Karina</name>

WinA. <date>06/01/2011</date>

Getaggt

<location>Krankenhaus der Samariter</location>

<location>Holzhausen</location>

Röntgenabteilung, Vorstand <name>Prim. Univ. Prof. Dr.Dr. Gotthard Vogler</name>

CT Abdomen und kl. Becken

Name: <name>Mustafa Üstün</name>, *
<date>21.06.67</date>

Aufnahmezahl: <id>1933309807</id>

Abteilung: Chirurgie

Station: <division>A31. OG. Viszeralchirurgie B</division> /

Zi: 119

dikt. Arzt: <name>OA Dr. Huber Karina</name>

WinA. <date>06/01/2011</date>

Anonymisiert

XXXXXXXXXX XXXXXXXXXXXX

Röntgenabteilung, Vorstand Prim. Univ. Prof. Dr.Dr. XXXXXXXX
XXXXXXXXXXXXXXXXXX

CT Abdomen und kl. Becken

Name: XXXXXXXX XXXXXXXXXXXX, * X.X.X

Aufnahmezahl: XXXXXXXXXXXXX

Abteilung: Chirurgie

Station: XXXXXXXXXXXX XXXXXXXXXXXXX

Zi: 119

dikt. Arzt: OA Dr. XXXXXXXXXXXX XXXXXXXXXXXX

WinA. XX/XX/XXXX

Getaggt

<location>Krankenhaus der Samariter</location>

<location>Holzhausen</location>

Röntgenabteilung, Vorstand <name>Prim. Univ. Prof. Dr.Dr. Gotthard Vogler</name>

CT Abdomen und kl. Becken

Name: <name>Mustafa Üstün</name>, *
<date>21.06.67</date>

Aufnahmezahl: <id>1933309807</id>

Abteilung: Chirurgie

Station: <division>A31. OG. Viszeralchirurgie B</division> /

Zi: 119

dikt. Arzt: <name>OA Dr. Huber Karina</name>

WinA. <date>06/01/2011</date>

Pseudonymisiert (separater Prozess)

Kantonsspital Friedrichshafen

Röntgenabteilung, Vorstand Prim. Univ. Prof. Dr.Dr.
Gerhard Voigtländer

CT Abdomen und kl. Becken

Name: **Manuel Überreuter**, * **1.07.69**

Aufnahmezahl: **9983209971**

Abteilung: Chirurgie

Station: **Station Sauerbruch**

Zi: 119

dikt. Arzt: OA Dr. **Heilmann Kristina**

WinA. **16/07/2013**

VERSCHIEBEN VON ZEITANGABEN UNTERSCHIEDLICHER GRANULARITÄT

Einheit	Algorithmus	Beispiel (offset = 300)
Tag	$d_{\text{pseudo}} = d_{\text{orig}} + \text{offset}$	11.03.2021 → 05.01.2022
Monat	$d_{\text{median}} = \text{floor} (\text{median} (\text{days} (m_{\text{orig}})))$ $m_{\text{pseudo}} = \text{month} (d_{\text{orig}} + \text{offset})$	3/2021 → 16.03.2021 16.03.2021 → 10.01.2022 → 1/2022
Jahr	$d_{\text{median}} = \text{floor} (\text{median}(\text{days}(y_{\text{orig}})))$ $y_{\text{pseudo}} = \text{year} (d_{\text{orig}} + \text{offset})$	2021 → 01.07.2021 01.07.2021 → 27.04.2022 → 2022
Nominale Zeitangaben		
Feiertage	Ersatz durch unspezifische Angaben oder Weglassen	"am Karfreitag, den 4. April operiert" -> "am Karfreitag, den 3. Januar" operiert -> "am 3. Januar operiert" "nach Pfingsten" -> "nach dem Feiertag"
Quartale, Jahreszeit	Analoges Vorgehen zu Monaten	Sommer 2021 → 07.08.2021 07.08.2021 → 03.06.2022 → Frühjahr 2022

PERSONENNAMEN: NORMALISIERUNG / KATEGORISIERUNG

Namensbestandteile		Beispiel
Trennzeichen	Leerzeichen	Anna Osler-> ('Anna', 'Osler')
Bindestrich	Kein Trennzeichen	Eva-Maria Rau ->('Eva-Maria','Rau')
1. Buchstabe klein	Ignoriert	'van Beethoven' -> 'Beethoven'
Token bis $\text{ceil}(\text{median}(n))$	Vornamen (V), dann Nachnamen (N)	'Kim Yong Il Park Un' V: 'Kim Yon Il', N: 'Park Un'
Namenszusätze		
Mit Punkt	Ignoriert	'Dr.', 'Prof.', 'Dipl.-Ing', 'Jr.'
Ohne Punkt	Ignoriert (aus Lexikon)	'OA', 'PD', 'PhD', 'MBA'
Repetitionen	Ignoriert	'Dr. Dr. Dr.' -> 'Dr'

PD Dr. med. Eva-Maria Gräfin von und zu Eulenhoven-Katzenfels PhD MBA

Namenszusatz
Vorname
Vorname
Ignoriere
Nachname
Namenszusatz

NAMEN: PSEUDONYMISIERUNG

Namen		Beispiel
Vornamen	Vornamenlexikon, indexiert nach Initiale und Geschlecht	Zufallszahl r $W = (\dots, 'Andrea', 'Agnes', \dots)$ 'Anna' $\rightarrow w_r$, z.B. 'Anna' \rightarrow 'Agnes'
Nachnamen	Nachnamenlexikon, indexiert nach Initiale	Zufallszahl r $N = (\dots, 'Emmerich', 'Emmersdorfer', 'Eils', \dots)$ 'Eberhard' $\rightarrow n_r$, z.B. 'Eberhard' \rightarrow 'Emmerich'
Kombinationen	Ausnahmen für seltene Initialien	'Yvonne Quervain' \rightarrow 'Annabel Müller'
Institutionen	Lexikon Pseudo-Lokalisationen Abgleich mit Terminologie	'LKH Salzburg' \rightarrow 'Scheuermann-Klinik' 'in HNO verlegt' \rightarrow 'in HNO verlegt'

PD Dr. med. Eva-Maria Gräfin von und zu Eulenhoven-Katzenfels PhD MBA


 PD Dr. med. Elisabeth Gundula Emmersdorfer PhD MBA