

# DSGVO konformer Umgang mit sensiblen Gesundheit- und personenbezogenen Daten in der Klinik

Dr. Frank Reddig, Leiter Medizincontrolling, Marienhospital Stuttgart

Volker Gertler, Regional Business Manager, 3M HIS

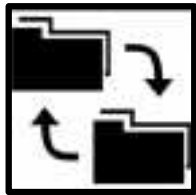
Krankenhausenerfolg durch Nutzen stiftende Digitalisierungsprojekte

# Agenda

- Hintergrund
- Ablauf / Lösung
- Praktische Erfahrung
- Projektskizzierung

*“Digitale Transformation kombiniert mit radikaler Interoperabilität von Daten ist das Schlüsselprinzip zur Generierung eines signifikanten Mehrwerts für Patienten, Behandler und das Gesundheitssystem.”*

# Datenaustausch – wichtig und komplex



Steigender Bedarf für Datenaustausch



Gesundheitsdaten sind besonders streng zu schützen



Die Datenschutzgrundverordnung regelt den Umgang mit Daten seit dem 25. Mai 2018 in Europa



Unstrukturierte Dokumente sind besonders schwierig zu schützen



# Unsere Lösung

Ursprungsdokument

Pseudonymisiertes /  
obfuskiertes Dokument



Sensitive Gesundheits-  
informationen werden  
intelligent ersetzt.

Obfuskation:  
Daten werden durch  
Daten vom gleichen Typ  
(Name durch Name)  
ersetzt.

Medizinische und  
kodierrelevante  
Informationen bleiben  
erhalten.

# Nutzen der pseudonymisierten Daten für Ihre Klinik

## Hausinterne Datensicherheit



**Interne Training-Systeme**  
Aus- und Fortbildung der Mitarbeitenden



**Compliance**  
Höhere Anforderungen von der europ. Datenschutzverordnung (EU-DSVO)



**Interner Datenzugriff**  
Limitierter Zugriff einzelner Abteilungen auf Original-daten, Peer-to-Peer Review

## Hausexterne Datensicherheit



**Externe Stellen**  
Datenaustausch mit haus-externen Stellen, Behörden, Auditoren, anderen Krankenhäusern



**Risiko-Reduktion als Unternehmung**  
Pseudonymisierte/ obfuskierte Dokumente reduzieren Re-Identifizierungsrisiko bei der Datenspeicherung



**Support Situationen**  
Für Ihren sicheren Datenaustausch mit Software-Partnern

## Forschung, Studien, Projekte



**Studien**  
Datenaustausch im Rahmen von wissenschaftlichen Studien



**Patienten-Rekrutierung**  
Datenaustausch im Rahmen der Patienten-Rekrutierung

## Heutiger Fokus



**Projektarbeiten**  
Datenaustausch im Rahmen von Projektarbeiten oder Tumorboards

# Das Marienhospital Stuttgart im Überblick



**2075**  
Mitarbeitende

**761**  
Planbetten

**Umsatz**  
>220  
Mio. €

Akademisches  
**Lehrkrankenhaus**  
der Uni Tübingen



Warum das MHS mittelfristig  
am Markt bestehen kann



81.500 ambulante  
33.500 stationäre  
Fälle pro Jahr  
(ohne Corona)



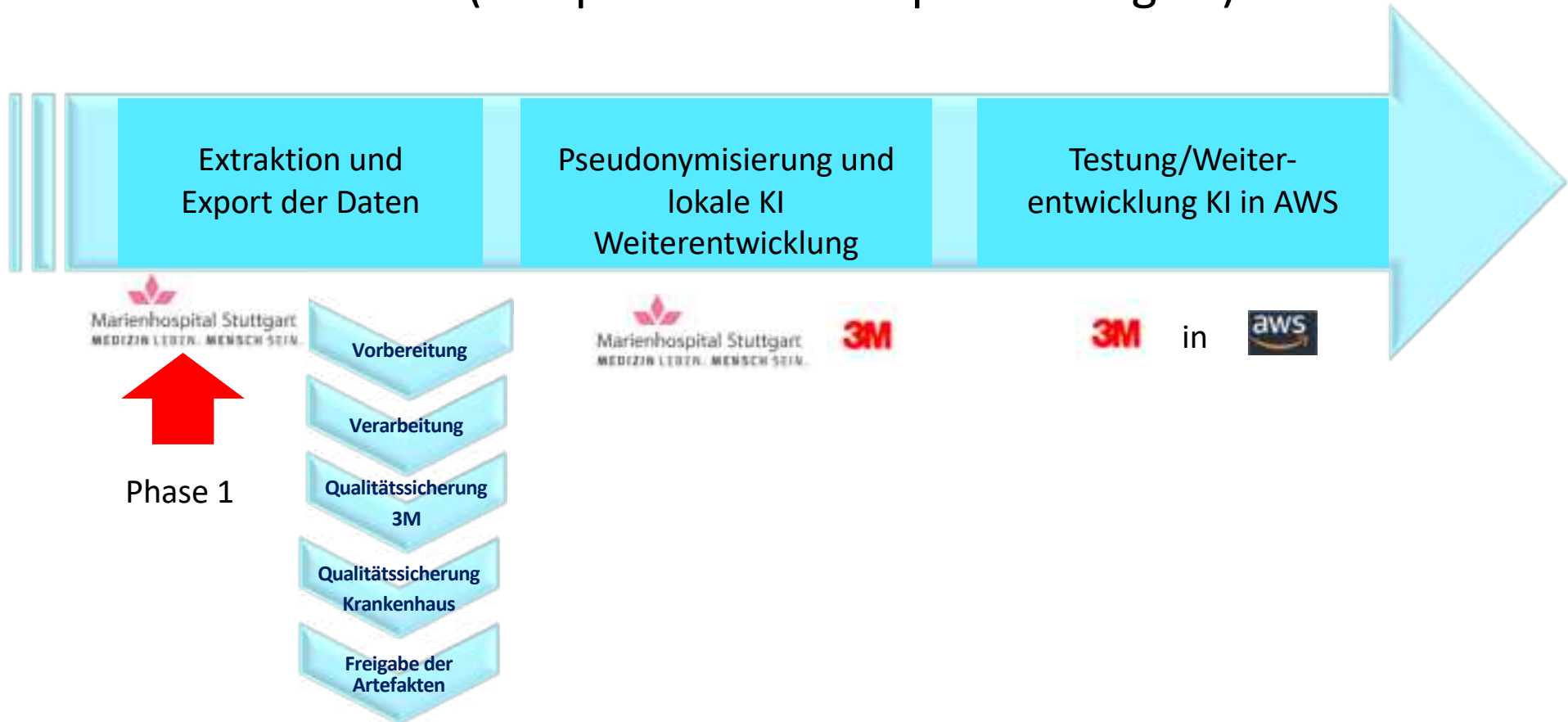
Eröffnet **1890**

18 Fachkliniken und  
19 interdisziplinäre Zentren



5,7 Tage  
Verweildauer

# Drei Phasen-Modell (Beispiel Marienhospital Stuttgart)







Erste Ergebnisse



# Dokumentenkollektion

Alle Dokumente 2020-21 wurde aus dem KIS DB als einfache Textdateien exportiert und umformatiert.

Insgesamt: **3.838.674** Dokumente von >86 Tsd Fälle ~ 50 Tsd Patienten)

Kodier-relevanz: **sehr hoch** – neben einer hohen Anzahl Arztbriefe und Konsiliarbefunde sind OP-Berichte und Patho-Befunde von allen Fachabteilungen des Hauses vorhanden.



# Analyse & Auswertungskriterien - Pseudonymisierung

Als Standard werden **5 Iterationen** durchgeführt.

- Nach jeder Iteration wird das Applikationslexikon mit neuen Termen und Mustern erweitert oder korrigiert.
- Jede Erweiterungen berücksichtigt die Wahrscheinlichkeit der Nutzung des Wortes im Kontext, Wortlänge, Lokalisation, Geschlecht.
- Die ursprüngliche Dokumentenkollektion wird nach jeder Iteration mit dem verbesserten Lexikon pseudonymisiert.

## Auswertung:

1. Wenn eine sensitive Information (Name, Vorname, Straße etc.) dem DeID-Tool unbekannt ist, oder als ein medizinisches Konzept erkannt wird, wird dies als **falsch negativ** interpretiert.
2. Wenn ein medizinisches Konzept dagegen als eine sensitive Information (Name, Vorname, Straße, etc.) erkannt wird, wird dies als **falsch positiv** interpretiert.
3. Unbekanntes Wort, als medizinisches Konzept erkannt und ein falscher Typ bei erkannten sensitiven Informationen werden ignoriert.

\*SHI – Sensitive Health Information

Ist-Typ	Soll-Typ	Falsch Negativ	Falsch Positiv
Unbekannt/Nicht-SHI	SHI	1	0
SHI	Nicht-SHI	0	1
Unbekannt	Nicht-SHI	0	0
SHI Typ X	SHI Typ Y	0	0



# Baseline Stichprobe – (N = 3400/3,8 Mil, CI = 98±2%)



TypeSHI	TotalSHI	%TotalSHI	%DistinctSHI	%AllTokens	Precision	Recall	FalseNegRate	FalsePosRate	F-Measure
UNKNOWN	1222	14,43	9,59	1,27	100	100	0	0	100
FNAME	1540	18,19	12,76	1,59	86,5	100	0	13,5	92,8
MI	4	0,05	0,14	0	100	100	0	0	100
LNAME	1956	23,1	22,17	2,03	99,9	99,8	0,2	0,05	99,8
PLACE	503	5,94	2,67	0,52	100	100	0	0	100
STREET	133	1,57	2,35	0,14	97,8	100	0	2,2	98,9
FACILITY	8	0,09	0,23	0,01	100	100	0	0	100
COUNTRY	1	0,01	0,05	0	100	100	0	0	100
EMPLOYER	8	0,09	0,14	0,01	100	100	0	0	100
IDNUM	889	10,5	16,13	0,92	100	100	0	0	100
ADDRNUM	51	0,6	1,24	0,05	100	100	0	0	100
DATE	2012	23,76	31,89	2,08	100	100	0	0	100
MONTH	22	0,26	0,41	0,02	100	100	0	0	100
YEAR	114	1,35	1,47	0,12	100	100	0	0	100
DAY	5	0,06	0,05	0,01	100	100	0	0	100
TOTAL-SHI	8468	100	100	8,77	96,9	99,9	0,05	3,1	98,4

\*SHI – Sensitive Health Information



# Qualitätssicherung nach Iteration 1/5 (800k Dokus) Stichprobe 1 (N = 3400/3,8 Mil, CI = 98±2%)



TypeSHI	TotalSHI	%TotalSHI	%DistinctSHI	%AllTokens	Precision	Recall	FalseNegRate	FalsePosRate	F-Measure
UNKNOWN	686	8,61	6,64	0,7	100	100	0	0	100
FNAME	1310	16,44	12,19	1,33	100	99,5	0,5	0	99,7
MI	15	0,19	0,27	0,02	100	100	0	0	100
LNAME	2051	25,74	22,15	2,08	100	99,2	0,8	0	99,6
PLACE	587	7,37	3,59	0,6	100	99,7	0,3	0	99,8
STREET	121	1,52	3	0,12	100	100	0	0	100
FACILITY	2	0,03	0,09	0	100	100	0	0	100
COUNTRY	1	0,01	0,05	0	100	50	50	0	66,6
EMPLOYER	6	0,08	0,14	0,01	100	100	0	0	100
TOTAL-SHI	7967	100	100	8,1	100	99,4	0,5	0	99,7

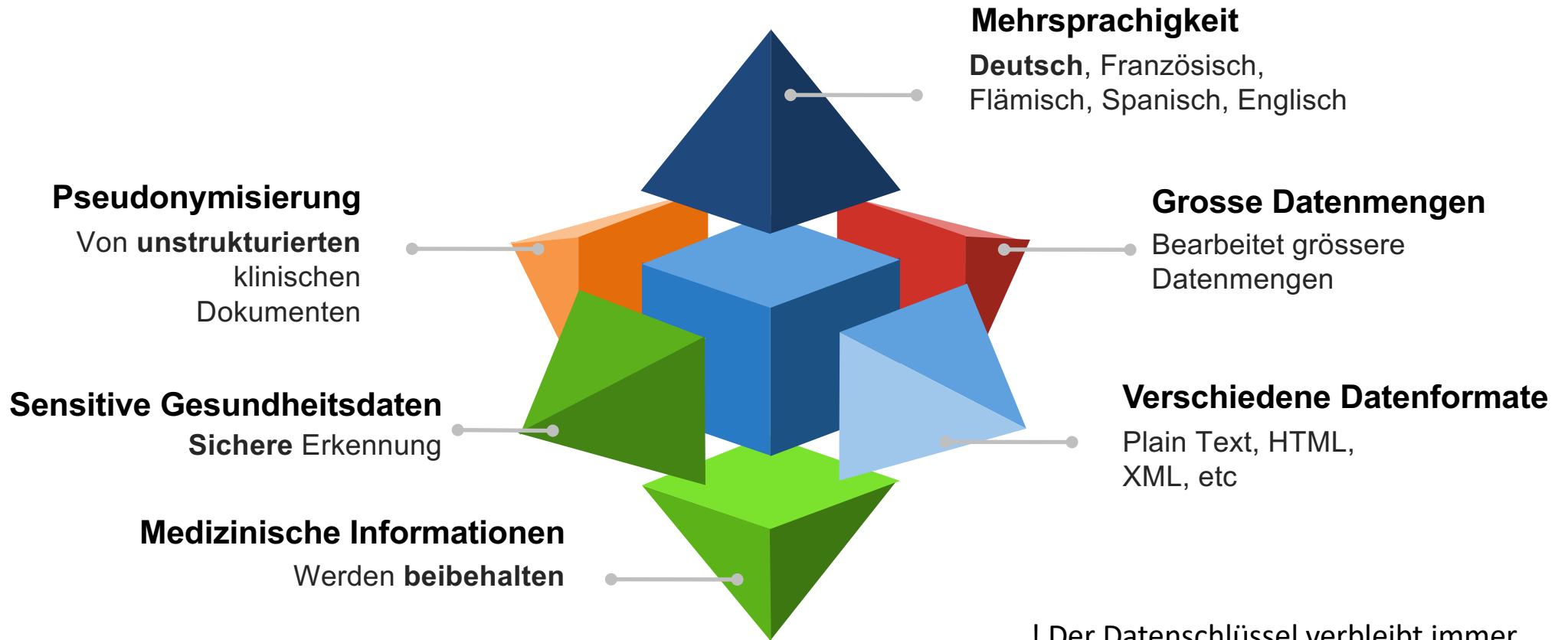
## Beispiel Krankenhaus nach 5 Iterationen (Jan. 2022)

	SHI Type	Total	Distinct	%Total	%Distinct	%AllTokens	Precision	Recall	FalseNegal	FalsePositi	F-Measure
Baseline	TOTAL-PHI	53246	4783	100	100	8.69	99.064	99.533	0.467	0.936	99.298
Status nach 5 Iterationen	TOTAL-PHI	151745	16094	100	100	10.12	99.624	99.999	0.001	0.376	99.811

\*SHI – Sensitive Health Information



# Details / Umfang der Lösung



! Der Datenschlüssel verbleibt immer  
in der Klinik!

## Projektskizzierung mit den potentiellen Teilnehmern

C  
H  
E  
C  
K

1. (a) Pseudonymisierung der eigenen Daten und enge Projektbegleitung durch 3M (Phase 1 des gezeigten Projektes)  
(b) Ziel: abschliessende Abnahme durch den Datenschutzbeauftragten des teilnehmenden Hauses -> KEINE DATEN VERLASSEN DAS HAUS!

K  
O  
N  
Z  
E  
P  
T  
E

2. Diskussion / Konzeptionierung / Nutzung der freigegebenen pseudonymisierten Daten für verschiedenste Anwendungen im operativen Umfeld, wie z.B.:
  - (a) Schulung / Training Personal
  - (b) Erhöhung der Hausinternen und -externen Datensicherheit
  - (c) Individuelle Anwendungen mit weiteren Partnern (Pharmaindustrie, EPA, MedTech..)
  - (d) weitere Anwendungsmöglichkeiten in Forschung und Lehre, wie z.B. Identifizierung von Patientengruppen
3. Gemeinsame Präsentation auf der MEDICA -> Folgeanwendungen identifizieren
4. NACHHALTIGKEIT(!): Die Teilnehmer erhalten nach Beendigung des Projektes für weitere 12 Monate kostenlos Updates und werden über Folgeanwendungen informiert.

Vielen Dank für Ihre Aufmerksamkeit Fragen/  
Anmerkungen?

DSGVO konformer Umgang mit sensiblen Gesundheit- und  
personenbezogenen Daten in der Klinik

Dr. Frank Reddig, Leiter Medizincontrolling, Marienhospital Stuttgart

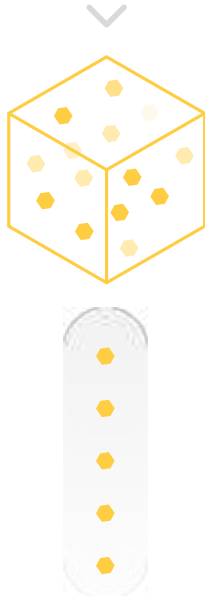
Volker Gertler, Regional Business Manager, 3M HIS



# Unsere Positionierung

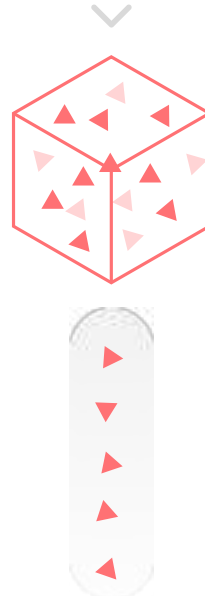
## Persönliche Information

Dokument mit persönlichen und Gesundheitsinformationen



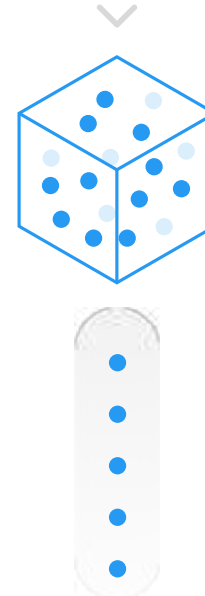
## Pseudonymisierung von Direct Identifiers

Direct Identifiers werden pseudonymisiert (analog zu den 18 HIPAA identifiers)



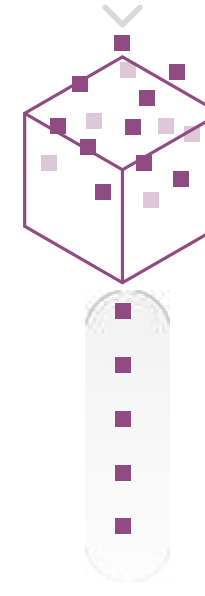
## Pseudonymisierung von weiteren Identifiers

Starke Pseudonymisierung mit direkten und indirekten Identifiers



## Anonymisierung

Anonymisierte Dokumente/Daten, document/data, irreversible Verarbeitung, keine Person identifizierbar



Innerhalb GDPR



Position 3M

Ausserhalb GDPR