SMHES
GMFH

Erfahrungen mit den gesicherten Daten des öffentlichen Gesundheitssystems in der Schweiz

Dr. David-Olivier Jaquet-Chiffelle

UNIL | Université de Lausanne
Faculté de droit
et des sciences criminelles

Berner Fachhochschule
Haute école spécialisée bernoise
Technik und Informatik
Technique et informatique

SMHES
GMFH

Beweggründe

- Pflegekosten in den Spitälern:
  17 Milliarden CHF im Jahre 1995
  (24 Milliarden CHF im Jahre 2006)
- Für welche Art von Pflege?
- Zu welchem Preis?
- Wo kann man sparen?

⇒ **Bedarf für eine Statistik**

SMHES
GMFH

Chronologischer Überblick

- VESKA (Verband Schweiz. Krankenanstalten)
  - Interne Statistik
  - Statistik mit Namen (keine Chiffrierung)
  - Während mehr als 20 Jahren
- BFS (Bundesamt für Statistik)
  - KVG (vom Parlament und Volk im Jahre 1994 gutgeheissen)
  - Bundesstatistikgesetz (Bundesrat)
  - Verordnung vom 30. Juni 1993 (über die Durchführung von statistischen Erhebungen des Bundes)

SMHES
GMFH

Bundesamt für Statistik

Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica
OFS BFS UST

Confidentiality and Data Protection - Patients Hospitalized in Switzerland

ISSE 2001, London
Electronic proceedings, Septembre 2001

J.-P. Jeanneret, D.-O. Jaquet-Chiffelle

SMHES
GMFH

Grundlegende Probleme

- Erfassung der Diagnose- und Behandlungsdaten für alle in der Schweiz hospitalisierten Personen
  Die Statistik ist **erschöpfend**

sensible Daten

- Notwendigkeit der Erkennung der Fälle von **Rehospitalisierung**
  - im gleichen Spital
  - in einem andern Spital

Verfolgung eines Patienten über mehrere Jahre hinweg

- Die **Anonymität** der Patienten garantieren

Datenschutz

SMHES
GMFH

Datentypen

- Identifizierende Daten
  - Name, Vorname
  - Geburtsdatum
  - Wohnort
  - etc.
- Epidemiologischen Daten
  - Alter (statt Geburtsdatum)
  - Region (statt Wohnort)
  - etc.

SMHES
GMFH

Triviallösung

„Wörterbuch“

Identifizierende Daten *(Name, Vorname, etc.)*	Persönlicher Code *(Eindeutige Zufallszahl)*
• Micheline Calmy-Rey	• 5723410846
• Pascal Couchepin	• 7291837102
• Doris Leuthard	• 3215639272
• Ueli Maurer	• 6232498237
• Moritz Leuenberger	• 1523612811
• ...	• ...

SMHES
GMFH

Triviallösung

„Zwei separate Datenbanken“

Datenbank der Personen

- Micheline Calmy-Rey 5723410846
- Pascal Couchepin 7291837102
- Doris Leuthard 3215639272
- Ueli Maurer 6232498237
- Moritz Leuenberger 1523612811
- ...

Datenbank der Behandlungen

- 7291837102
  - Behandlung 1, in BE, am 18/2/2003
  - Behandlung 2, in ZH, am 22/6/2007
  - etc.
- 1523612811
  - Behandlung 1, in LU, am 28/3/2004
  - Behandlung 2, in ZH, am 16/5/2006
  - etc.
- ...

Schlecht angepasst an die gesetzlichen Rahmenbedingungen in der Schweiz und zu verletzbar

SMHES
GMFH

Berechneter persönlicher Code

Gewünschte Eigenschaften:

- Die identifizierenden Daten gestatten die einfache Berechnung des persönlichen Codes.
- Der persönliche Code gestattet nicht die identifizierenden Daten wiederzufinden.
- Die gleiche Person erhält immer den gleichen persönlichen Code.
- Zwei verschiedene Personen erhalten fast immer verschiedene persönliche Codes. Das „fast immer“ erzeugt in der Datenbank ein leichtes Rauschen.
  - dies erhöht noch das Niveau des Datenschutzes
  - ohne die statistischen Untersuchungen zu beeinträchtigen

SMHES
GMFH

Kryptographie

- Gesetzlicher Rahmen in der Schweiz
  - Keine Einschränkung
- Gesetzlicher Rahmen in Frankreich
  - Vor 1999
    - Genehmigung notwendig
  - Seit 1999 (Dekret Nr. 99-200 vom 17. März 1999)
    - Gebrauch von kryptographischen Algorithmen frei von jeglichen Formalitäten, falls die Schlüssellänge weniger als 40 Bits umfasst
  - Nach 2004 (Gesetz über das Vertrauen in die digitale Wirtschaft Nr. 2004-575 vom 21. Juni 2004)
    - Der Gebrauch von kryptographischen Mitteln ist frei (Art. 30 I)

SMHES
GMFH

Minimale identifizierende Daten

- Gewünschte Eigenschaften:
  - unterscheidend
  - unabhängig vom Spital
  - immer verfügbar
  - zeitlich konstant
- Praktische Wahl: (Kompromiss)
  - Familienname, Vorname, Geburtsdatum, Geschlecht

Maximal 5'000*5'000*(365,25*120)*2 = 2'191'500'000'000 Möglichkeiten (Entropie ist kleiner als 41)

SMHES
GMFH

Identifikationsnummer im Register in Frankreich (NIR)

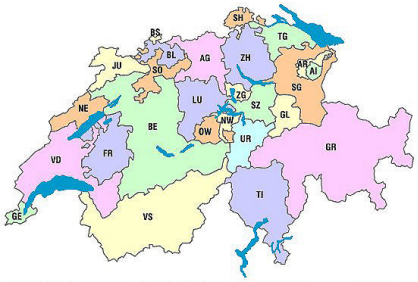
Fall	Position	Bedeutung	Mögliche Werte	Anzahl Möglichkeiten		
alle	1	Geschlecht: 1 für Männer, 2 für Frauen	1 oder 2			2
	2, 3	die zwei letzten Ziffern des Geburtsjahres	von 00 bis 99			100
	4, 5	Geburtsmonat	von 01 bis 12, oder 20			13
A	6, 7	Geburtsdepartement im französisches Mutterland (2A oder 2B für Korsika)	von 01 bis 95, 2A, 2B	97	96'030	98'820
	8, 9, 10	Nummer der Geburtsgemeinde im Departement	von 001 bis 990	990		
B	6, 7, 8	Geburtsdepartement in einem französischen Überseegebiet	von 970 bis 989	20	1'800	
	9, 10	Nummer der Geburtsgemeinde im Departement	von 01 bis 90	90		
C	6, 7	Geburtsort ausserhalb von Frankreich	99	1	990	
	8, 9, 10	Landeskennnummer des Geburtslandes	von 001 bis 990	990		
alle	11, 12, 13	Nummer der Geburtsurkunde im Monat und in der Gemeinde (oder des Landes)	von 001 bis 999	999		999
	14, 15	*Kontrollschlüssel modulo 97*	*von 01 bis 97*	*1*		*1*

Maximal 256'675'068'000 Möglichkeiten (Entropie kleiner als 38)

SMHES
GMFH

Erste Transformation (T1)

- Falls möglich sollte die Transformation **T1** nicht von einem geheimen Schlüssel abhängen.



⇒ Kryptographische Einweg-Hashfunktion ohne Schlüssel

SMHES
GMFH

Rechtschreibefehler

- **Jaquet-Chiffelle**
- Jaquet-Chiffele
- Jaquet-Chifelle
- Jaquet-Chifele
- Jaquet-Schiffelle
- Jaquet-Schiffele
- Jaquet-Schifelle
- Jaquet-Schifele

- Jacquet-Chiffelle
- Jacquet-Chiffele
- Jacquet-Chifelle
- Jacquet-Chifele
- Jacquet-Schiffelle
- Jacquet-Schiffele
- Jacquet-Schifelle
- Jacquet-Schifele

SMHES
GMFH

Robuste Transformation

- Nicht zwischen Gross- und Kleinbuchstaben unterscheiden
- Elimination der Akzente, der Leerschläge, der Bindestriche
- Ersetzung von “y” durch “i”
- Komprimierung der Doppelkonsonanten
  - “ff” wird “f”
  - “ss” wird “s”, etc.
- Ersetzung von “sch” oder “sh” durch “ch”
- Ersetzung von
  - “ae” durch “a”
  - “ou” oder “ue” durch “u”
- Elimination der “h” , ausser sie folgen auf ein “c”
- Ersetzung von “cqu” durch “qu”
- etc.

SMHES
GMFH

Berücksichtigte Wahl für T1

- Berücksichtigte robuste Transformation der identifizierenden Daten
  - Soundex-Code des Familiennamens
  - Soundex-Code des Vornamens
- SHA-1 (**S**ecure **H**ash **A**lgorithm)
  - hängt nicht von einem geheimen Schlüssel ab
- Kompression
  - Abdruck von 64 Bit

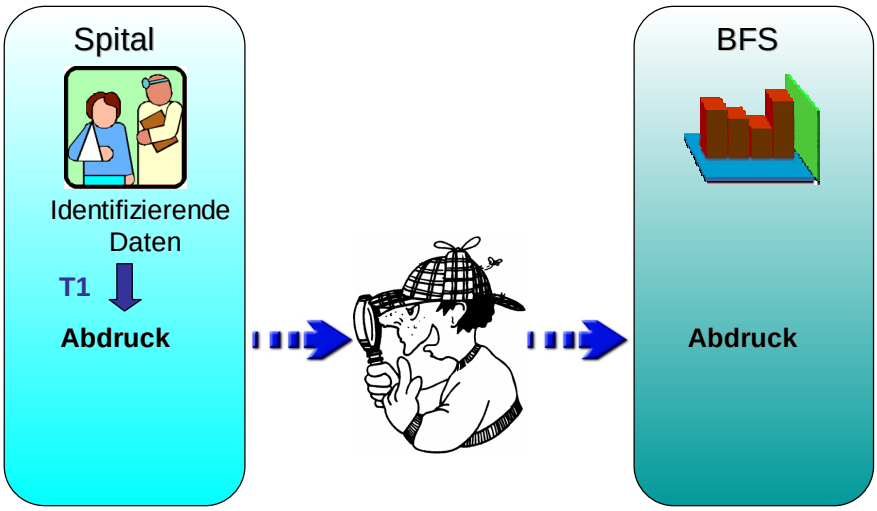
SMHES
GMFH

Validierung (reale Daten)

- Realer Test mit der Datenbank der **Genfer Universitätsspitäler (HUG)** (Dr. Borst)
  - 222'000 Einträge
  - Kollisionsrate **< 0.3%**
  - Aufspürung von Doppeldaten
- Positiver Effekt: Fehlerbereinigung der Datenbank der Genfer Universitätsspitäler

SMHES
GMFH

Übertragung ans BFS



SMHES
GMFH

Verwundbarkeit von T1

- Die Transformation **T1** sollte nicht von einem geheimen Schlüssel abhängen.
- Da **T1** nicht von einem geheimen Schlüssel abhängt ist der Abdruck verwundbar.
  - Punktueller Angriff
  - Wörterbuchangriff
- Der Abdruck **verbirgt** nur die Identität.

➢ Notwendigkeit der Chiffrierung der Abdrücke

SMHES
GMFH

Verschlüsselung des Abdrucks

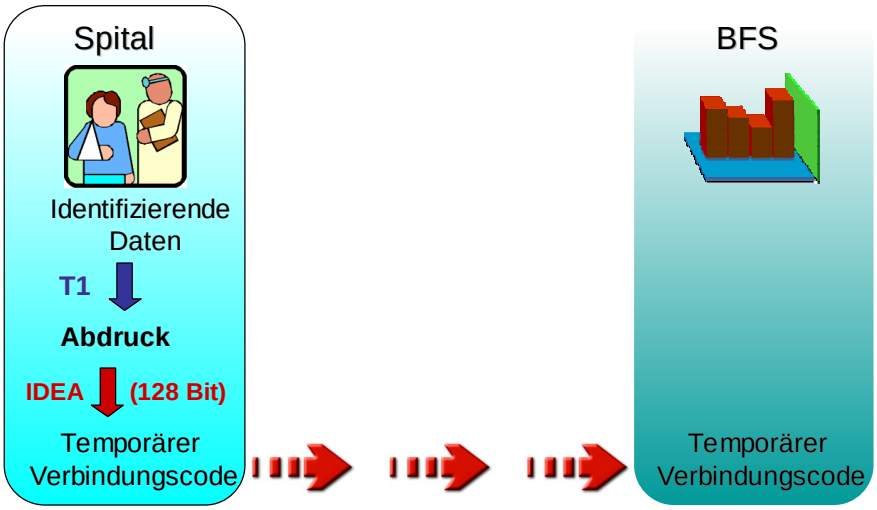
IDEA

International **D**ata **E**ncryption **A**lgorithm

- Symmetrischer Verschlüsselungsalgorithmus
- Geheimer **Sitzungsschlüssel** von 128 Bit
- Eingang und Ausgang: Blöcke von 64 Bit

SMHES
GMFH

Temporärer Verbindungscode



Windisch, 19. Juni 2009 © 2009, Jaquet-Chiffelle
Seite 21

SMHES
GMFH

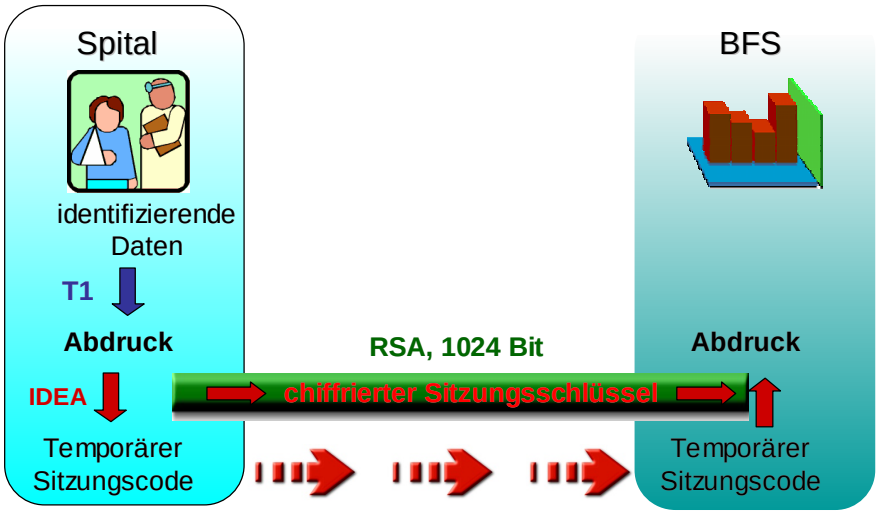
Sitzungsschlüssel

- Werden von den Spitälern verwendet, um die Abdrücke während der Übertragung ans BFS zu verschlüsseln (Temporäre Verbindungscodes)
  - IDEA, Sitzungsschlüssel von 128 Bit
- Übertragung ans BFS mit Hilfe eines Public-Key-Chiffrierverfahrens
  - RSA, 1024 Bit
- Gestattet dem BFS die Abdrücke wiederzufinden, um die gleichmässigen anonymen Verbindungscodes zu berechnen

Windisch, 19. Juni 2009 © 2009, Jaquet-Chiffelle
Seite 22

SMHES
GMFH

Public-Key-Kryptographie



SMHES
GMFH

Berücksichtigte Wahl für das Pendant von T1 (in Frankreich)

- Anonymisierende Funktion FOIN *(Fonction d'Occultation des Identifiants Nominatifs)*
- Basierend auf
  - der Sozialversicherungsnummer
  - Geburtsdatum des Patienten
  - Geschlecht des Patienten
- Hash-Funktion mit einem fixen geheimen Schlüssel
- Geteilter geheimer Schlüssel, wird aber an alle Spitäler verteilt

SMHES
GMFH

Zweite Transformation (T2)

- Vom BFS zur Erzeugung des **anonymen (gleichmässigen) Verbindungscodes** verwendet, die den **berechneten persönlichen Code bestimmt**
- Muss von einem geheimen Schlüssel K abhängen

Zur Wahl:

- Hash-Funktion mit geheimen Schlüssel (F)
  - "FOIN 2" in Frankreich
- Chiffrieralgorithmus (CH)

SMHES
GMFH

Berücksichtigte Wahl für T2 in der Schweiz

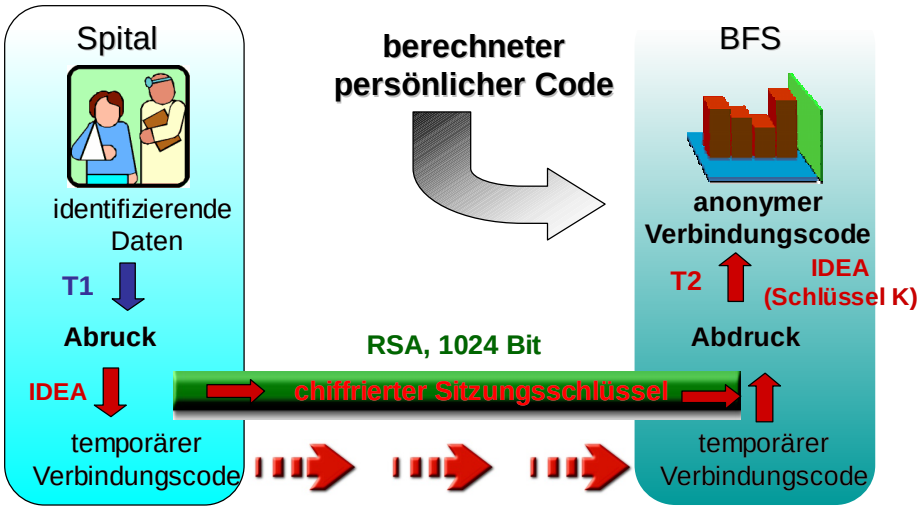
IDEA

International **D**ata **E**ncryption **A**lgorithm

- Symmetrischer Chiffrieralgorithmus
- Geheimer Schlüssel K von 128 Bit
- Eingang und Ausgang: Blöcke von 64 Bit

SMHES
GMFH

Anonymer Verbindungscode



SMHES
GMFH

Secret Sharing

- Der private RSA-Schlüssel des BFS (der für die Decodierung der temporären Verbindungscodes verwendet wird) und
- der geheime Schlüssel K des BFS (der für die Berechnung der gleichmässigen, anonymen Verbindungscodes ausgehend von den Abdrücken verwendet wird, das heisst für die Berechnung der persönlichen Codes)

stellen zusammen ein sensibles Geheimnis dar.

In unserem Protokoll, wird dieses Geheimnis aufgeteilt.

SMHES
GMFH

Vertrauenspersonen

- Präsident des Schweizerischen Ärzteverbandes
- Direktor des BFS
- Der eidgenössische Datenschutz- und Öffentlichkeitsbeauftragte (EDÖB)

SMHES
GMFH

Secret Sharing

1001011011100010 11

- schlechte Strategie...

1001011011100010 11

- gute Methode:

Teile

1001011011100010 11	geheim
0011001101001011 01	zufällig
0100111010011001 00	zufällig
1110101100110000 10	XOR

SMHES
GMFH

Aktuelle Situation

- Alle Spitäler in der Schweiz haben ihre krypto-graphischen Module homologiert (eigentlich seit 1999)
- Alle Spitäler in der Schweiz (gegenwärtig 321 Spitäler) benutzen dieses System um ihre Daten an das BFS zu übertragen
- Andere Organisationen interessieren sich für dieses System (Beispiele: Versicherungen, Eidg. Justiz- und Polizeidepartement, etc.).

SMHES
GMFH

Verwendung in der Schweiz

- 1998: erste Verwendung des Systems im grossen Massstab (ungefähr 38% aller Hospitalisierungsfälle).
- 2006: Das System hat es den Spitälern erlaubt 1'240'678 Fälle zu übermitteln (erschöpfende Statistik, 100% der Fälle).
  - Diese 1'240'678 Fälle betreffen nur 899'454 Patienten.

SMHES
GMFH

Statistische Verteilung (2006)

Mehrfachhospitalisierungen in der Schweiz
Jahr 2006 (erschöpfende Statistik)

Anzahl Tage	Anzahl Patienten	Rate der Mehrfach-Hospitalisierungen	Anzahl Fälle
1	692'252	77.8%	692'252
2	136'489		272'978
3	40'412		121'236
4	15'964		63'856
5	6'869		34'345
6	3'358	22,2%	20'148
7	1'755		12'285
8	943		7'544
9	519		4'671
10	311		3'110
11+	582		8'256
Total	**889'454**	**100%**	**1'240'678**
Ausgeschlossene Fälle (schlecht kodierte Patienten): « 0000 »			127

Publikation der Daten im März 2008

SMHES
GMFH

Verwendung in der Schweiz

- 1998: Erste Verwendung des Systems im grossen Massstab ungefähr 38% aller Hospitalisierungsfälle).
- 2006: Das System hat es den Spitäler gestattet 1'240'678 Fälle zu übertragen (erschöpfende Statistik, 100% der Fälle).
  - Diese 1'240'678 Fälle betreffen nur 899'454 Patienten.
- 2007: Das System hat es den Spitälern gestattet 1'279'918 Fälle zu übertragen (erschöpfende Statistik, 100% der Fälle).
  - Diese 1'279'918 Fälle betreffen nur 923'529 Patienten.

SMHES
GMFH

Statistische Verteilung (2007)

Mehrfachhospitalisierungen
Im Jahre 2007 (erschöpfende Statistik)

Anzahl Tage	Anzahl Patienten	Rate der Mehrfach-Hospitalisierungen	Anzahl Fälle
1	708'282	76.7%	708'282
2	141'047		282'094
3	42'330		126'990
4	16'670		66'680
5	7'192		35'960
6	3'600	23,3%	21'600
7	1'886		13'202
8	1'043		8'344
9	497		4'473
10	345		3'450
11+	637		8'843
Total	**923'529**	**100%**	**1'279'918**
Ausgeschlossene Fälle (schlecht codierte Patienten): « 0000 »			112

Publikation der Daten im März 2009

SMHES
GMFH

Verwendung in der Schweiz

- 1998: Erste Verwendung des Systems im grossen Massstab (ungefähr 38% aller Hospitalisierungsfälle).
- 2006: Das System hat es den Spitäler gestattet 1'240'678 Fälle zu übertragen (erschöpfende Statistik, 100% der Fälle).
  - Diese 1'240'678 Fälle betreffen nur 899'454 Patienten.
- 2007: Das System hat es den Spitälern gestattet 1'279'918 Fälle zu übertragen (erschöpfende Statistik, 100% der Fälle).
  - Diese 1'279'918 Fälle betreffen nur 923'529 Patienten.
- Die anonymen Verbindungscodes gestatten eine sehr präzise Beschreibung der Verteilung der Mehrfach-hospitalisierungen ohne die Identität der Patienten zu enthüllen.

SMHES
GMFH

Schlussfolgerung

- Der anonyme Verbindungscode
  - gestattet die Erkennung der Fälle von Mehrfach-hospitalisierungen
  - garantiert die Anonymität der Patienten mit einem hohen Sicherheitsniveau
  - erzeugt ein leichtes Rauschen in den Daten
    - was der Schutz der Daten noch erhöht
    - ohne die statistischen Studien zu beeinträchtigen
  - ist anpassbar an ein vergleichbares Umfeld (Versicherungen, Rechtssysteme, etc.)
- Alle Algorithmen sind öffentlich

SMHES
GMFH

Fragen?

UNIL | Université de Lausanne
Faculté de droit
et des sciences criminelles

Berner Fachhochschule
Haute école spécialisée bernoise
Technik und Informatik
Technique et informatique

david-olivier.jaquet-chiffelle@bfh.ch